

The construction of facial composites by witnesses with mild learning disabilities

By
Julie Gawrylowicz


A thesis submitted for the degree of
Doctor of Philosophy
awarded by the University of Abertay Dundee

October 2010

The submitted thesis is in partial fulfilment of the requirements of the University of Abertay Dundee for the award sought.

This thesis is a copy of the true and accurate version of that approved by the examiners,..... (signature principal supervisor).....
(Date).

Acknowledgments

First of all, I would like to thank my **supervisors**  Dr Derek Carson and Dr Fiona Gabbert for their advice, support and encouragement during my PhD and beyond. I am very grateful for their patience, effort and time they spent to help me with my studies, thesis, publication and job applications. Special thanks to Fiona for not only acting as a supervisor but also as a friend in various situations.

Many thanks also to my former MSc supervisors Professor Amina Memon and Dr Ingrid Candel, who encouraged me to pursue a PhD in the first place.

I would also like to thank several people at the University of Abertay who assisted me during my PhD in several ways. Without their help it would have taken me much longer to complete my studies and it would not have been such a pleasant experience. Thanks to Donald Williamson, Elaine Whyte, Eric Turnbull, Pamela Cord, Ros Elliott, Dawn Keen, and Mary Wilkie.

A big ‘thank you’ goes to all the service users and members of staff from the day care centres in and around the Dundee area. Without their help I would not have been able to successfully complete these studies. Special thanks go to Steve Hamilton, Brian Herd, and Andrea Crawford for their continuous help and motivation.

Thanks to my friends and colleagues Sonja, Mo, Kat, and Sharon for all the experiences, coffee breaks, and glasses of wine we shared and for keeping me sane.

I am especially thankful to my boyfriend Helge for being my partner on our PhD rollercoaster rides. Thanks for the frequent lifts to the train- and bus station, for your understanding, inspiration and care. Thanks for being you!

My biggest thank goes to Mämchen, Pöpchen and Oli for their unconditional love and emotional as well as financial support. Thanks for letting me whinge, laugh and cry about my PhD and for always providing a place of refuge.

Finally, I would like to thank the University of Abertay Dundee and the Scottish Institute for Policing Research for generously funding the research.

Abstract

In a criminal investigation, witnesses may get asked by the police to provide a perpetrator's description or to generate a composite image of the perpetrator's face. Due to their elevated vulnerability to victimisation people with a learning disability (LD) may be more likely than other members of the wider community to find themselves in such situations. Research regarding face recognition and description abilities of this group has been to some extent neglected in the eyewitness research literature. Consequently, guidance for practitioners on how to effectively generate facial composite images with LD witnesses is limited. The current research addresses this issue, by investigating basic and applied face recognition and description abilities in individuals with mild learning disabilities (mLD) during a series of experimental studies. Moreover, potential facilitating measures are introduced and assessed.

Five studies were conducted during the course of this thesis. In the first study a survey was designed to collect information on currently used composite systems by UK law enforcement agencies and how operators perceive and treat witnesses with LD. The survey findings confirmed the initial assumption that individuals with LD may indeed find themselves in the situation of having to describe a perpetrator's face to an investigative officer. Furthermore, the results emphasised the lack of guidance available to operators on how to best meet the special needs of this particular witness population.

Study 2 investigated basic face recognition and description abilities in people with mLD and revealed that overall they performed at a lower level than the non-LD controls. Despite this finding, mLD individuals as a group performed above chance levels and they displayed variability in performance depending on the introduced measures.

Studies 3 and 5 investigated these abilities in a more applied setting, namely during the construction of facial composites with contemporary facial composite systems. Study 3 revealed that composites generated with the E-FIT system, a featural system, were considerably poorer than those created by their non-LD counterparts. Studies 4 and 5 attempted to improve mLD individuals' performance by applying visual prompts and by using a more holistic facial composite system, i.e. EvoFIT. There was little evidence of the former being advantageous for witnesses with mLD, however, EvoFIT significantly enhanced composite construction abilities in the mLD participants.

Finally, the practical and theoretical implications of the main findings are discussed.

List of Figures

		Page
4.1:	Bruce and Young's (1986) functional model for human face recognition	62
4.2:	Two-dimensional representation of the exemplar based face-space model by Valentine (1991)	71
6.1:	Graph displaying percentage of facial composite systems operators have experience with	113
6.2:	Graph displaying percentage of facial composite systems operators are currently working with	114
6.3:	Graph showing the different sources which provided training to operators	115
7.1:	Picture of practice task 1 used during Experiment 1 in Study 2	142
7.2:	Picture of practice task 2 used during Experiment 1 in Study 2	143
7.3:	Picture of main recognition task used during Experiment 1 in Study 2	144
7.4:	Graph depicting the mean number of facial information (+SE) provided by mLD- and control participants during the different description modes (photo vs. memory) and recall conditions (free vs. cued) in Experiment 2, Study2	155
7.5:	Graph showing the mean percentages of appropriate facial information (+SE) provided by mLD- and control participants during the different description modes (photo vs. memory) and recall conditions (free vs. cued) in Experiment 2, Study 2	158
7.6:	Picture depicting the matching task used in the evaluation phase in Study 2	162
7.7:	Graph showing the mean percentages of correct identifications (+SE) for facial descriptions provided by mLD and control participants during the different description modes (photo vs. memory) in Experiment 2, Study 2	164
8.1:	Picture of a screenshot of the E-FIT system	176
8.2:	Graph displaying the significant interaction between program and group obtained in Study 3	185
8.3:	Graph showing the mean number of percentages of accepted exemplars (+SE) by the two groups of participants during the two different description modes while creating facial composites with E-FIT in Study 3	191
8.4:	Picture showing an example of the matching task in Study 3	194
8.5:	Graph showing the mean number of correct matches (+SE) for E-FITs created by mLD and control participants during the different description modes during the evaluation phase of Study 3	195

8.6:	Picture showing an example of the likeness rating task in Study 3	197
8.7:	Graph depicting the mean likeness rating scores (+SE) for composites created by mLD and control participants during the different description modes (photo vs. memory) obtained during the evaluation phase in Study 3	198
9.1:	Picture showing the two target faces used during the visual prompt task in Study 4	208
9.2:	Picture displaying the visual prompt for eyebrow size with the three different prompt options: thin, medium-sized or thick.	209
10.1:	Example of face triads used during the easy (left side) and difficult (right side) morph tasks in Study 5	226
10.2:	Picture of screenshot of the selection and breeding phase during the composite construction procedure with EvoFIT in Study 5	234
10.3:	Picture of screenshot of the holistic tool phase during the composite construction procedure with EvoFIT in Study 5	235
10.4:	Picture of screenshot of the shape tool phase during the composite construction procedure with EvoFIT in Study 5	235
10.5:	Graph showing the mean number of facial features requested to be changed (+SE) for both participant groups during the two different description modes in Study 5	241
10.6:	Picture of the matching task used in Study 5	244
10.7:	Graph depicting the mean number of correct matches (+SE) obtained during the matching task in Study 5	245
10.8:	Picture of the likeness rating task in Study 5	247
10.9:	Graph depicting the mean likeness rating scores (+SE) obtained during Study 5	248
10.10:	Graph depicting the significant interaction between group and program in Study 5	254
10.11:	Graph showing the significant three-way interaction between group, program and description mode obtained in Study 5	256

List of Tables

	Page
7.1: Overview of findings obtained during Experiment 1 of Study 2.	150
7.2: Inter-coder reliability correlations calculated during Experiment 2 in Study 2	153
7.3: Mean number of facial information (+SD) provided by mLD and control participants during the different description modes (photo vs. memory) and recall conditions (free vs. cued) during the description task in Experiment 2, Study 2	153
7.4: Mean number of facial information (+SD) provided by mLD- and control participants during the different description modes (photo vs. memory) and recall conditions (free vs. cued) during Experiment 2 in Study 2	156
7.5: Mean percentages of accurate facial information (+SD) provided by mLD and control participants during the different description modes (photo vs. memory) and recall conditions (free vs. cued) in Experiment 2, Study 2	156
7.6: Mean percentages (+SD) of correct identifications for facial descriptions provided by mLD and control participants during the different description modes (photo vs. memory) obtained during the evaluation phase in Experiment 2, Study 2	163
8.1: Mean number of facial information (+SD) provided by mLD and control participants during the CIs for the different description modes in Study 3	179
8.2: Mean number of times mLD and control participants pointed to their own faces (+SD) during the two description modes in Study 3	181
8.3: Mean number of times mLD and control participants pointed to the monitor (+SD) during the two description mode conditions in Study 3	182
8.4: Mean durations (+SD) spent in the E-FIT program and the Paint program for both participant groups during Study 3	184
8.5: Mean number of facial features (out of seven) requested to be changed (+SD) for both participant groups during the two different description modes during the facial composite construction with E-FIT in Study 3	186
8.6: Mean number of external and internal facial features requested to be changed (+SD) by the two participant groups during the different description modes during the composite construction with E-FIT in Study 3	188
8.7: Mean number of exemplars participants requested to see before being satisfied with one (+SD) during the two description modes through the composite construction with E-FIT in Study 3	189
9.1: Mean number of prompt options (out of 24) (+SD) selected by mLD and control participants which were consistent with the scoring sheet.	212

9.2:	Correlations revealed between mLD individuals' WASI scores and the amount of selected prompt options which were consistent with the scoring sheet.	215
9.3:	The most frequently selected prompt options by control participants for target face 1 during the visual prompt task in Study 4	216
9.4:	The most frequently selected prompt options by control participants for target face 2 during the visual prompt task in Study 4	217
9.5:	The most frequently selected prompt options by mLD participants for target face 1 during the visual prompt task in Study 4	218
9.6:	The most frequently selected prompt options by mLD participants for target face 2 during the visual prompt task in Study 4	218
9.7:	Display of prompts, for which the selected prompt option was consistent between mLD and control participants for target faces 1 & 2, respectively, together with the level of agreement in percentages within each participant group in Study 4	219
10.1:	Mean number of correct responses during the easy and difficult morph tasks (+SD) by mLD and control participants during Study 5	228
10.2:	Correlations obtained between mLD participants scores on the WASI and the morph tasks in Study 5	230
10.3:	Mean number of correct responses (hits) (+SD) during the practice tasks (easy and difficult) by mLD and control participants in Study 5	238
10.4:	Mean durations of time (+SD) spent in the three different composite construction phases for both participant groups and during the two description modes in Study 5	238
10.5:	Correlations obtained between participants WASI scores, performance on practice task, durations of composite construction phases, number of changed features with shape tool and the composite quality during Study 5	250
10.6:	Correlations obtained between participants performance on practice task, durations of composite construction phases, number of changed features with shape tool and the composite quality during Study 5	251
10.7:	Mean number of correct matches and likeness rating scores (+SD) obtained on the basis of E-FIT and EvoFIT composites created by either mLD or control participants during the two different description modes in Study 5	252

List of Appendices

		Page
1	A copy of the survey of police operators used in Study 1	305
2	Male facial stimuli used during the description task in Study 2	320
3	Facial stimuli used during the E-FIT composite construction in Study 3	321
4	The complete set of visual prompts used during the visual prompt task in Study 4	322

List of Abbreviations

ACPO	Association of Chief Police Officers
AMRE	Attitudes towards Mental Retardation and Eugenics
ARC	Adult Resource Centre
CA	chronological age
CCQ	Credibility Comparisons Questionnaire
CI	Cognitive Interview
CJS	Criminal Justice System
EA	evolutionary algorithm
ECI	Enhanced Cognitive Interview
E-FIT	Electronic Facial Identification Technique
EvoFIT	Evolutionary Facial Identification Technique
FSIQ	Full Scale IQ
K-S test	Kolmogorov-Smirnov test
LD	learning disability
MA	mental age
mLD	mild learning disability
MOPI	management of police information
NPIA	National Policing Improvement agency
PACE	Police and Criminal Evidence Act
PCA	principal component analysis
PITO	Police Information Technology Organisation
RT	reaction time
SCLD	Scottish Consortium for Learning Disability
SI	Standard Interview
TA	target absent
TP	target present
VOE	verbal overshadowing effect
WAIS	Wechsler Adult Intelligence Scale
WASI	Wechsler Abbreviated Scale of Intelligence

Contents

ACKNOWLEDGMENTS	II
ABSTRACT.....	III
LIST OF FIGURES	V
LIST OF TABLES	VII
LIST OF APPENDICES	IX
LIST OF ABBREVIATIONS	X
CONTENTS.....	XI
CHAPTER 1.....	15
GENERAL OVERVIEW, AIMS AND STRUCTURE OF THESIS	15
1.1 GENERAL OVERVIEW.....	15
1.2 THESIS STRUCTURE	18
1.2.1 <i>General overview of thesis</i>	18
1.2.2 <i>Literature review chapters</i>	19
1.2.3 <i>Experimental studies</i>	20
1.2.4 <i>General discussion</i>	21
CHAPTER 2.....	22
DEFINITION, PREVALENCE AND VICTIMISATION RATES OF LD	22
2.3 DEFINING LD	22
2.4 PREVALENCE RATES OF PEOPLE WITH LD	25
2.2.1 <i>Prevalence rates of people with LD in the UK</i>	25
2.2.2 <i>Prevalence rates of people with LD in England</i>	25
2.2.3 <i>Prevalence rates of people with LD in Scotland</i>	27
2.2.4 <i>Prevalence rates of people with LD in Wales</i>	27
2.2.5 <i>Summary of prevalence rates</i>	27
2.3 VICTIMISATION RATES OF PEOPLE WITH LD	28
2.3.1 <i>Barriers to obtain information</i>	29
2.3.2 <i>Reasons for higher victimisation</i>	30
CHAPTER 3.....	32
THE PERFORMANCE OF PEOPLE WITH LD AS EYEWITNESSES.....	32
3.1 INTRODUCTION.....	32
3.2 ATTITUDES TOWARDS WITNESSES WITH LD.....	33
3.2.1 <i>Police officers' attitudes towards individuals with LD</i>	34
3.2.2 <i>Jurors' attitudes towards witnesses with LD</i>	34
3.2.3 <i>Treatment of witnesses with LD in court by lawyers and judges</i>	35
3.3 RESEARCH REGARDING EYEWITNESS ACCOUNTS OF INDIVIDUALS WITH LD	37
3.3.1 <i>Eyewitness accounts by children with LD</i>	37
3.3.2 <i>Eyewitness accounts by adults with LD</i>	42
3.4 WAYS TO IMPROVE EYEWITNESS ACCOUNTS OF LD INDIVIDUALS.....	44
3.4.1 <i>The Cognitive Interview</i>	44
3.4.2 <i>Repeated questioning of individuals with LD</i>	49
3.5 SPECIAL LEGAL MEASURES FOR WITNESSES WITH LD	50
3.6 DISCUSSION	54
CHAPTER 4.....	56
REVIEW OF APPLIED RESEARCH IN FACE RECOGNITION.....	56

AND RECALL	56
4.1 INTRODUCTION.....	56
4.2 FACE RECOGNITION.....	59
4.2.1 <i>Familiar versus unfamiliar face recognition</i>	59
4.2.2 <i>Situational factors influencing face recognition</i>	63
4.2.3 <i>Face specific factors</i>	68
4.3 FACE RECALL.....	76
4.3.1 <i>Verbal overshadowing effect</i>	77
4.3.2 <i>Feature saliency</i>	80
4.3.3 <i>Forensic implications</i>	81
4.4 GENERAL CONCLUSIONS	82
CHAPTER 5.....	84
HISTORY AND DEVELOPMENT OF FACIAL COMPOSITE SYSTEMS	84
5.1 WHAT IS A FACIAL COMPOSITE?	84
5.2 EARLY FACIAL COMPOSITE SYSTEMS.....	85
5.2.1 <i>Sketch artist</i>	85
5.2.2 <i>Mechanical composite systems</i>	86
5.3 CURRENT FACIAL COMPOSITE SYSTEMS	91
5.3.1 <i>Computer-based composite systems</i>	92
5.3.2 <i>Evolutionary composite systems</i>	99
5.4 DISCUSSION	106
CHAPTER 6.....	109
STUDY 1: A SURVEY OF FACIAL COMPOSITE OPERATORS	109
6.1 INTRODUCTION.....	109
6.2 METHOD	110
6.2.1 <i>Participants</i>	110
6.2.2 <i>Survey instrument and procedure</i>	111
6.3 RESULTS	112
6.3.1 <i>Personal details</i>	112
6.3.2 <i>General experiences with and opinions about composite construction</i>	117
6.3.3 <i>Experiences with and opinions about witnesses with LD</i>	120
6.4 DISCUSSION	129
6.4.1 <i>Current practices utilised by police operators during the composite construction</i> .	129
6.4.2 <i>Operators' experiences with and attitudes towards witnesses with LD</i>	131
6.4.3 <i>Practical problems operators' face during the composite construction with LD</i> <i>witnesses</i>	133
6.4.4 <i>Limitations of the survey study</i>	134
CHAPTER 7.....	136
STUDY 2: FACE RECOGNITION AND DESCRIPTION ABILITIES IN PEOPLE WITH MLD OF UNFAMILIAR FACES.....	136
7.1 INTRODUCTION.....	136
7.1.1 <i>Research aims of Study 2</i>	138
7.2 EXPERIMENT 1 FACE RECOGNITION SKILLS IN PEOPLE WITH MLD	139
7.2.1 <i>Method</i>	140
7.2.2 <i>Results</i>	144
7.3 EXPERIMENT 2 FACE DESCRIPTION SKILLS IN PEOPLE WITH MLD	150
7.3.1 <i>Phase 1 Describing unfamiliar faces</i>	150
7.3.2 <i>Phase 2 Evaluation of facial descriptions</i>	160
7.4 DISCUSSION	165
CHAPTER 8.....	171
STUDY 3: THE EFFICIENCY OF E-FIT WITH MLD WITNESSES	171
8.1 INTRODUCTION.....	171

8.2	PHASE 1: COMPOSITE CONSTRUCTION	173
8.2.1	<i>Method</i>	173
8.2.2	<i>Results</i>	178
8.3	PHASE 2: COMPOSITE EVALUATION	193
8.3.1	<i>Matching task</i>	193
8.3.2	<i>Likeness rating task</i>	196
8.3.3	<i>Factors influencing composite quality</i>	198
8.3.4	<i>Summary of main findings</i>	200
8.4	DISCUSSION	201
CHAPTER 9		204
STUDY 4: DO VISUAL PROMPTS FACILITATE VERBAL DESCRIPTIONS OF UNFAMILIAR FACES IN WITNESSES WITH MLD?		204
9.1	INTRODUCTION	204
9.1.1	<i>Research aims</i>	206
9.2	METHOD	207
9.2.1	<i>Design</i>	207
9.2.2	<i>Participants</i>	207
9.2.3	<i>Materials</i>	208
9.2.4	<i>Procedure</i>	209
9.2.5	<i>Scoring</i>	211
9.3	RESULTS	211
9.3.1	<i>Analysis of data with a more stringent cut-off criterion</i>	213
9.3.2	<i>Correlations between performance and WASI scores</i>	214
9.3.3	<i>In depth exploration of the raw data</i>	215
9.4	DISCUSSION	220
CHAPTER 10		222
STUDY 5: THE SUITABILITY OF EVOFIT FOR WITNESSES WITH MLD		222
10.1	INTRODUCTION	222
10.2	PART 1: THE MORPH TASK	224
10.2.1	<i>Method</i>	225
10.2.2	<i>Results</i>	227
10.3	PART 2: COMPOSITE CONSTRUCTION	231
10.3.1	<i>Method</i>	231
10.3.2	<i>Results</i>	236
10.4	PART 3: COMPOSITE EVALUATION	243
10.4.1	<i>Matching task</i>	243
10.4.2	<i>Likeness rating task</i>	246
10.5	EVOFIT VS. E-FIT	252
10.5.1	<i>Matching task</i>	252
10.5.2	<i>Likeness rating task</i>	254
10.5.3	<i>Summary of main findings</i>	256
10.6	DISCUSSION	257
CHAPTER 11		267
GENERAL DISCUSSION AND CONCLUDING REMARKS		267
11.1	INTRODUCTION	267
11.2	SUMMARY AND MAIN FINDINGS	268
11.2.1	<i>A survey of facial composite operators</i>	268
11.2.2	<i>Face recognition and description abilities in people with mLD of unfamiliar faces</i>	269
11.2.3	<i>The efficiency of E-FIT with mLD witnesses</i>	269
11.2.4	<i>Do visual prompts facilitate verbal descriptions of unfamiliar faces in witnesses with mLD?</i>	270
11.2.5	<i>The suitability of EvoFIT for mild learning disabled witnesses</i>	270
11.3	DISCUSSION OF MAIN FINDINGS	271
11.3.1	<i>The ability of people with mLD to recognise and verbally describe faces</i>	272

11.3.2	<i>The ability of people with mLD to construct facial composites</i>	277
11.3.3	<i>Individual differences in face recognition and description</i>	284
11.4	METHODOLOGICAL LIMITATIONS	287
11.5	CONCLUDING REMARKS	291
REFERENCES		293
APPENDICES		305

Chapter 1

General overview, aims and structure of thesis

1.1 General overview

Victims of crime and witnesses to crimes where the perpetrator is unknown are often required to provide a facial description of the perpetrator's face to the police. This verbal facial description can subsequently be used to create a facial composite image. This image can be circulated to other police services or the general public to assist in the search for the offender. The quality of the verbal description and the resulting composite image can play a crucial role in the successful outcome of a criminal investigation.

It is estimated that the UK has a high prevalence rate of people with learning disabilities (LD), with 2.5% of the population falling into this category (BILD, 2006). Emerson (2001) argues that this figure is likely to rise in the future. There are reasons to believe that people with LD are more likely to find themselves in the situation of having to describe the face of a perpetrator to the police. Possible reasons for this include the fact that people with LD often live in underprivileged neighbourhoods with high crime rates (Hatton & Emerson, 1996), and are frequently used by criminals to assist in illegal activities (Davies, 1995). Moreover, only a small proportion of cases involving victims with LD are reported to authorities (Valentin-Hein & Schwartz, 1993), this may leave those individuals even more vulnerable to future victimisation because perpetrators are less likely to be frightened of retribution (Milne & Bull, 2006). All these factors and others might contribute to the situation, where

people with LD are disproportionately represented in the criminal justice system as victims or witnesses.

On the other hand, people with LD are often excluded from normal criminal justice procedures, such as creating a facial composite of the perpetrators' face, because they are regarded by a significant proportion of people as less credible and accurate witnesses (Peled, Iarocci & Connolly, 2004; Stobbs & Kebbell, 2003). In a recent Scottish crime case, a woman with LD was raped and assaulted by several men. No prosecution emerged primarily because she was considered an unreliable witness by The Crown Office (Severin, 2008). In a similar case, reported by O'Hara (2001), a woman suffering from severe LD had been sexually abused by a family member. Despite available evidence, the case never reached court, because according to the law, her disability made her incapable of giving consent in court. The ACPO (Association of Chief Police Officers) Working Group for Facial Identification stated in 2003 that "Serious considerations should be given to the potential evidential value and accuracy of the recognition and recall factors from: very young or old witnesses, witnesses who are mentally impaired, and witnesses impaired by alcohol or drugs." (ACPO, 2003, p.10). These examples clearly show that, in the absence of relevant forensic evidence, individuals with LD are more likely than their non LD counterparts to be excluded from general criminal justice procedures.

Despite the scepticism regarding the ability of witnesses with LD to provide reliable evidence it is surprising to find that research on their ability to provide verbal descriptions or to construct facial composites of unfamiliar faces has been somewhat neglected.

On the contrary, a fair amount of research has been conducted in the eyewitness domain, concerning the accuracy and completeness of recall of an

observed event provided by people with LD (Agnew & Powell, 2004; Brown & Geiselman, 1990; Cederbrog, La Rooy & Lamb, 2008; Michel, Gordon, Ornstein & Simpson, 2000; Milne, 1999; Milne, Clare & Bull, 1999; Robinson & Mc Guire, 2006). Most of those studies investigated the performance of children with LD (Agnew & Powell, 2004; Brown & Geiselman, 1990; Cederbrog, La Rooy & Lamb, 2008; Milne & Bull, 1996) with only a few studies investigating the performance of adults (Brown & Geiselman, 1990; Milne, Clare & Bull, 1999; Perlman, Ericson, Esses & Isaac 1994). A consistently reported research finding is that people with LD are more suggestible than control participants and they show a higher tendency of acquiescence (Gudjonsson & Henry, 2003; Henry & Gudjonsson, 1999; Hery & Gudjonsson, 2003; Milne, Clare & Bull, 2002). Moreover, it has been established that the reliability of information provided by LD individuals is highly influenced by the question format utilised (Agnew & Powell, 2004). Nevertheless, overall research findings suggest that witnesses with LD are able to give accurate accounts of observed events when questioned in an appropriate way taking their disability into account (see Chapter 3 for a detailed review of the impact of question format on mLD witnesses' accounts). The question arises whether this also holds for the description of unfamiliar faces and the creation of facial composite images.

The process of creating a facial composite of an unfamiliar face involves numerous cognitive components, such as verbal description, recall and recognition of individual facial features (Pike, Kemp & Brace, 2000). Individuals with LD often have limited cognitive abilities, involving deficits in memory, language comprehension and production and decision making (Swanson, Cooney & O'Shaughnessy, 2004). These cognitive deficiencies might act as a barrier to

accurately describe and recognise previously seen faces and to engage efficiently in the construction of facial composite images.

The aim of this thesis is to fully investigate the ability of people with mLD to recognise and describe unfamiliar faces, and to use facial composite systems, such as E-FIT (Electronic Facial Identification Technique) and EvoFIT (Evolutionary Facial Identification Technique) to construct accurate facial composite images. The first part of the thesis reviews the available literature regarding eyewitness performance in individuals with LD, relevant applied face recognition studies, and recall research, and the development of facial composite systems. The experimental section provides a detailed examination of face recognition and description abilities in witnesses with mLD, thereby filling the gap in the existing research literature. The findings from the thesis provide an insight into the difficulties that people with mLD might experience when using facial composite software and suggest ways how they might be assisted. The overall purpose of the thesis is to improve current strategies used by the police to meet the special needs of people with mLD. Furthermore, the results of this research may be used to help write guidelines for police officers, advocates, and judges on how to obtain best evidence from witnesses with mLD. It is hoped that the impact of the findings will help contribute to a fairer and more reasonable treatment of individuals with LD by the criminal justice system.

1.2 Thesis structure

1.2.1 General overview of thesis

- *Chapter 1: General overview, aims and structure of thesis*

1.2.2 Literature review chapters

The literature review chapters highlight all aspects addressed in the current thesis. They include definitions and prevalence rates of LD, research regarding the ability of people with LD as eyewitnesses and theoretical and practical aspects concerning the production of facial composites.

- *Chapter 2: Definition, prevalence and victimisation rates of LD*

Chapter 2 outlines the different definitions of LD and specifies which criteria are used during the experimental studies of this thesis to select participants with mLD. To further highlight the relevance of the research project, current and future hypothesised prevalence rates of people with LD in the UK are presented as are victimisation rates and contributing factors to these rates.

- *Chapter 3: The performance of people with LD as eyewitnesses*

Chapter 3 gives an overview of past and current research regarding the performance of people with LD as eyewitnesses. The chapter covers topics such as the impact of different question formats on the accuracy and reliability of recall, suggestibility, and special measures for LD witnesses during legal procedures.

- *Chapter 4: Review of applied research in face recognition and recall*

Chapter 4 provides a concise overview of previous and contemporary research regarding humans' abilities to recall and recognise faces and describes influential theoretical frameworks. Relevant insights regarding face recognition are addressed and the limited research on face recall is reviewed.

- *Chapter 5: History and development of facial composite systems*

Chapter 5 provides an overview of the historical development of facial composite systems, and research evaluating these systems is discussed. Specific emphasis is put

on composite systems, such as E-FIT and EvoFIT, since these are the two composite systems utilized during the experimental studies.

1.2.3 Experimental studies

The main body of the thesis describes novel empirical work. The first study is a survey study, during which police operators' opinions about, and experiences with, LD witnesses are explored. The following studies investigate general face recognition and description abilities in mLD individuals. Thereafter, their performance with facial composite systems, such as E-FIT and EvoFIT is tested and attempts at improving this performance are outlined.

- *Chapter 6: Study 1 A survey of facial composite operators*

The first study involves a survey of UK police operators exploring current police service usage of facial composite systems and how operators treat and perceive witnesses with LD.

- *Chapter 7: Study 2 Face recognition and description abilities in people with mLD of unfamiliar faces*

The aim of the second study is to compare the ability of people with mLD and control participants to recognise and describe faces. The experiment includes three old/new face recognition tasks and two face description tasks. The results suggest that there is initial evidence that people with mLD are consistently poorer in performance on face recognition and recall tasks, fitting with the generally held layman's view that they might be less reliable eyewitnesses. However, there is also evidence that these individuals exhibit variability in performance dependent on the task. This suggests that they might benefit from measures introduced to facilitate performance.

- *Chapter 8: Study 3 The efficiency of E-FIT with mLD witnesses*

Having established basic face recognition and description abilities in participants with mLD, the third study investigates the performance of mLD participants in a more applied setting, namely during the construction of facial composites with E-FIT. Composites are constructed on the basis of the facial descriptions provided by the participants, and are subsequently evaluated by an independent sample of participants using a matching task and a likeness rating task.

- *Chapter 9: Study 4 Do visual prompts facilitate verbal descriptions of unfamiliar faces in witnesses with mLD?*

Study 4 investigates the effectiveness of visual prompts as a potential tool to assist participants with mLD to accurately describe unfamiliar faces.

- *Chapter 10: Study 5 The suitability of EvoFIT for mLD witnesses*

During the 5th Study, the ability of mLD participants to use a more holistic approach to facial composites, i.e. EvoFIT, is examined. The relative paucity of information provided by witnesses with mLD and the fact that EvoFIT does not require a verbal facial description, might make this system more suitable for witnesses with mLD.

1.2.4 General discussion

- *Chapter 11: General discussion and concluding remarks*

The final chapter comprises a general discussion of the findings and concluding remarks.

Chapter 2

Definition, prevalence and victimisation rates of LD

This chapter outlines different definitions of LD and summarises what they all have in common. On the basis of these similarities, criteria are determined which are used during the remainder of this thesis to select participants with mLD. Following this, current and future hypothesised prevalence rates of people with LD in the UK are presented as are victimisation rates.

2.3 Defining LD

The term ‘learning disability’ was adopted by the UK Government in 1991 (Welsh Assembly Government, 2007). It is a controversial term, with different definitions being utilized depending on the country, the domain (legal, educational, social) and the date of reference. Although there has been an abundance of research concentrating on this population in recent years no uniform definition has been established that holds general acceptance (Ashton & Ward, 1992; Hogg, 2001). As a result of this, and in an attempt to avoid any confusion, it is important to define at the outset the criteria utilised throughout the current experimental studies when referring to participants with mLD.

The term ‘learning disability’ replaced earlier terms, like ‘mental deficiency’ and ‘mental handicap’, which were frequently used during the 1960s and 1970s (Hogg, 2001). The main impetus for such changes was that the former terms were regarded as negative and discriminating. The term ‘learning disability’ no longer includes the stigmatization of being ‘mental’ and therefore gained acceptance by several sources and also by those to whom the previous terms had been applied

(Hogg, 2001). In the United States of America (USA), the terms ‘mental retardation’ and ‘developmental disability’ are still quite common, however, internationally the term ‘intellectual disability’ is more widely used (Emerson, Hatton, Felce & Murphy, 2001). In the UK the term ‘learning disability’ is utilized by most services, carers and professionals (Emerson et al., 2001).

Several definitions of LD have been put forward by the World Health Organization (WHO), the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), and The British Psychological Society (BPS). In 1992 the WHO published the tenth revision of the International Classification of Diseases (ICD-10) which states that LD (mental retardation) is defined as “...a condition of arrested or incomplete development of the mind, which is especially characterized by impairment of skills manifested during the developmental period, which contribute to the overall level of intelligence, i.e. cognitive, language, motor, and social abilities.” (p. 176). According to the DSM-IV (1995), LD (mental retardation) is characterized by “(a) significantly sub average intellectual functioning: an IQ of approximately 70 or below on an individually administered IQ test, (b) Concurrent deficits or impairments in present adaptive functioning (i.e. the persons’ effectiveness in meeting the standards expected for his or her age by his or her cultural group) in at least two of the following areas: communication, self-care, home-living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health, and safety, and (c) The onset is before age 18 years” (p. 49). The British Psychological Society (BPS) (2000) recognised that irrespective of the specific terminology, the different definitions have three core criteria in common, these are: significant impairment of intellectual functioning, significant impairment of adaptive/social functioning and the

age of onset is before adulthood. All three criteria must be met for an individual to be considered as learning disabled.

It should be recognised that people with LD do not constitute a homogeneous group. There might be several individual differences in severity, extent and nature of specific restrictions on performance caused by the LD (BPS, 2000). The WHO (1992) divided individuals with LD into four subcategories: mild, moderate, severe, and profound. The division is based on IQ score which is usually assessed with the Wechsler Adult Intelligence Scale (WAIS, 1981). The WAIS is a reliable and widely used instrument to measure intellectual functioning. Mild LD refers to an IQ score of 50 to 70, moderate LD refers to an IQ score of 35 to 49, severe LD refers to an IQ score of 20 to 34 and profound LD refers to an IQ score of 20 or below.

To summarise, there are specific elements which constitute LD that have attained general acceptance within the UK. First of all, LD is characterized by a significant impairment of intellectual functioning, second, there is also a significant impairment of adaptive and social functioning present, and finally, the disability must have been acquired before the age of 18 years. These three criteria were also utilized during the following studies to identify participants with mLD. All participants with mLD had an IQ score of 50 to 70, which lies in the classification range of mLD according to the WHO (1992). Furthermore, all participants received assistance in their daily routines, since they were all service users of Adult Resource Centres (ARC) in Scotland. Finally, it was confirmed by members of staff of the ARCs that the onset of the mLD was before adulthood in the participants.

2.4 Prevalence rates of people with LD

2.2.1 Prevalence rates of people with LD in the UK

It is difficult to estimate the prevalence rate of people with LD, since differences across studies in classification criteria, assessment methods, geography, language and culture make direct comparisons of studies nearly impossible (Emerson et al., 2001). Nonetheless, available statistics suggest that a significant number of people in the UK have LD and that this amount is likely to increase in the future. Emerson et al. (2001) attempted to estimate UK prevalence rates by combining rates from studies across North America, Europe and Australia and a 1995 estimate of the UK population (58.3 million). The results suggested that in the UK between 230,000 and 350,000 people have severe LD and between 580,000 and 1,750,000 people have mLD. The high prevalence rate is the most apparent reason why research regarding the eyewitness performance of people with LD is highly relevant.

2.2.2 Prevalence rates of people with LD in England

The most recent review considering prevalence rates of LD in England was conducted by Emerson and Hatton in 2008. The review requested by the Department of Health estimated that 985,000 people in England had LD, which corresponds to 2% of the general population. This number included 828,000 adults (aged 18 or above), of whom 174,000 were 60 years or older, and 174,000 people that made use of LD services. Earlier, the Department of Health (2001) estimated that 210,000 people in England suffer from severe and profound LD, including approximately 65,000 children and young people, 120,000 adults and 25,000 older people. In addition, approximately 1.2 million English adults have mild or moderate LD (2.5% of the general population). Interestingly, the estimate for adults with mild and moderate LD

by the Department of Health is higher than the estimate given by Emerson and Hatton (2008). This might be due to the fact that the Department of Health used the same prevalence estimate (2.5%) across all age groups, whereas Emerson and Hatton (2008) only used this prevalence estimate for the 15 to 24 years age group but lower estimates for older age groups because of the increased mortality rates.

Emerson and Hatton (2008) argue that, whatever the current rate, there will in the future be a likely increase. This increase might be due to two factors: general demographic changes in the English population and specific changes in the incidence and prevalence of people suffering from LD. Current population predictions suggest that the English population will increase from 50.9 million in 2007 to 53.5 million in 2017 and 56.0 million in 2027. These changes in the general population of England will probably result in equivalent changes in the population of people with LD. Therefore, it is likely that the overall amount of people with LD in England will rise proportionally. When considering specific changes in the proportion of the population with LD, Emerson and Hatton (2008) pointed out three factors which might lead to an increase in these figures over the next two decades. First of all, there is an increase in the prevalence of younger English adults belonging to Bangladeshi and Pakistani South Asian minority ethnic communities and evidence suggests that there might be a two or three fold increase in the prevalence of more severe and profound LD in children and younger adults belonging to these ethnic populations. Secondly, there is an increased survival rate of young people suffering from severe and profound LD. And finally, there is a higher life expectancy of people with LD in general. Moreover, the Department of Health (2001) mentioned that the number of reported cases of school age children with autistic spectrum disorders is very likely to rise in the future.

Given the fact that some of those children will have LD, the number of reported cases of children suffering from LD is likely to increase as well.

2.2.3 Prevalence rates of people with LD in Scotland

The most comprehensive data available with respect to the population of adults with LD in Scotland is provided by eSAY (2008). The Scottish Consortium for Learning Disability (SCLD) runs eSAY. It comprises a database, which includes national information about learning disability and autistic spectrum disorder in Scotland. Their annual statistical report covers the adult population with LD that is known to local authorities. In 2008, the estimated amount of people with LD in Scotland amounted to 25,252, which corresponds to approximately six adults with LD per 1,000 population. Furthermore, the most recent report stated that the overall number of people suffering from LD known to local authorities in Scotland increased by 40% between 2003 and 2008.

2.2.4 Prevalence rates of people with LD in Wales

There is less literature available regarding prevalence rates of people with LD in Wales. The Welsh Assembly Government (2007) reported that in 2006, 13,422 adults with LD were known to local authorities, a rate of 4.5 per 1,000 of the Welsh population. Unfortunately, the Welsh Assembly Government gives only estimates for the Welsh population suffering from profound LD. Thus, no further statistics are available.

2.2.5 Summary of prevalence rates

Overall, these numbers suggest that the prevalence rate of people suffering from LD is high in the UK with research indicating a likely increase in the future. This highlights

the relevance of conducting research regarding their performance as eyewitnesses. However, it should be noted, that these statistics might be an underestimation of the real number of individuals with LD, since most studies have used the number of LD people known to local authorities and it is likely that not all cases are reported to them. Care should be also taken when considering studies that have generalised prevalence rates from different geographical areas, because it might be that differences in methodology and terminology might have affected the results.

2.3 Victimisation rates of people with LD

As mentioned in the general introduction, people with LD might be more likely than their non LD counterparts to find themselves in the situation of having to describe the face of a perpetrator to the police because of their higher vulnerability to victimisation. Studies of victimisation started in the middle of the 20th century (Petersilia, 2001). It was soon realized that four groups of people are at a significantly higher risk to victimisation than other people of the general population; these are children, the elderly, females and the disabled (Petersilia, 2001). An early literature review carried out by Sobsey and Varnhagen (1988) reported that people with disabilities are at a greater risk of suffering sexual abuse. However, this report did not differentiate between cognitive and physical disabilities, making it difficult to generalise their findings. In an Australian study conducted by Wilson and Brewer (1992), a victimisation questionnaire was administered to adults with LD, ranging from mild to severe LD. The research showed that people with LD are twice as likely to be victims of crime directed against them personally, and they are one and a half times more likely to experience property crimes than individuals without LD.

Moreover, Wilson and Brewer found that the relative risk of victimisation was highest for personal crimes, e.g. sexual assault, assault, robbery, etc.

The literature shows that crimes committed against people with LD are quite diverse, including murder, violent personal assault, personal and property crime, financial victimisation and exploitation by overpersuasive sales techniques (Nettelbeck & Wilson, 2002). Another category of crime which specifically affects people with LD are so called 'hate crimes', which imply violence against LD individuals motivated by prejudices and the perception of them as being vulnerable targets (Nettelbeck & Wilson, 2002). Taken together, this suggests that people with LD are disproportionately at risk of being victims of certain kinds of crime. Furthermore, these individuals might be more vulnerable to recruitment into cult membership and for making false confessions under police interrogation (Nettelbeck & Wilson, 2002). Although, there seems to be a general agreement in the literature that people with LD are at a higher risk of victimisation, most literature lacks scientific research, which might be the result of weak methodology and difficulties to obtain relevant information (Nettelbeck & Wilson, 2002; Petersilia, 2001).

2.3.1 Barriers to obtain information

Why is it so difficult to obtain data on the victimisation of LD individuals? First of all, the majority of crimes involving victims with disabilities are not reported to the police. This might be due to problems in communicating the incident as well as anxiety to report the crime, because the victim might have a dependent relationship to the perpetrator; he or she might be a carer or a family member (Petersilia, 2001). Wilson and Brewer (1992) found that the rates of reporting crimes to the police were considerably low. Forty percent of the crimes against people with mLD and up to 71% of crimes committed against people with severe LD were not reported. The

situation becomes even more problematic due to the circumstance that when crimes against people with disabilities are reported, they are often handled as cases of abuse or neglect rather than crimes, and hence are dealt with internally rather than becoming public and being investigated by the police (Nettelbeck & Wilson, 2002; Petersilia, 2001). Another important factor which possibly influences the fact that a lot of cases are not reported involves the attitudes of police officers towards people with LD. Since police officers are among the first people who encounter information from victims and witnesses to crimes, they might also be the ones making decisions about the credibility of the informant and whether the case will be investigated further. A survey study conducted by Bailey, Barr and Bunting (2001) revealed that police officers possessed certain 'eugenic attitudes' (p. 348) towards individuals with LD. This might have an impact on whether crimes involving witnesses or victims with LD will be taken seriously and investigated further.

2.3.2 Reasons for higher victimisation

Three main factors are described in the literature which might be responsible for the high victimisation rate in people with LD. Firstly, people with LD are more likely to live in underprivileged neighbourhoods with high crime rates in general (Hatton & Emerson, 1996). Due to a shift of less institutionalised forms of care, such as long-stay hospitals and residential housing in the community, more people with LD are living independently or in their family home. This might have advantages as well as disadvantages. People with LD might enjoy a more active presence in the community and live more independent lives, but on the other hand living in the community might increase the risk of victimisation, since they receive less additional support from social services and the health care system (Nettelbeck & Wilson, 2002; Petersilia, 2001; VOICE, 2001). Secondly, people with LD are often used by criminals to assist

in illegal activities because of their low understanding of their actual involvement in a crime and their heightened need to be accepted by other people (Davies, 1995). Thirdly, victims with LD might not have the courage to report the crime to authorities, because of their communication difficulties, or they might have the impression that nobody will believe them. Some people might also experience feelings of shame or guilt which might prevent them from reporting the criminal incident (VOICE, 2001). Moreover, offenders are seldom successfully prosecuted due to a lack of evidence and credibility of the victim (Agnew & Powell, 2004). This makes individuals with LD even more vulnerable to future victimisation because perpetrators might not be frightened of retribution (Milne & Bull, 2006).

In conclusion, the high prevalence rate of people with LD in the UK and the fact that they are more vulnerable to victimisation than other members of the general public emphasise the importance of conducting more research regarding their performance as eyewitnesses. Specifically, research is needed to investigate their abilities to recognise, describe and construct facial composites of unfamiliar faces since this area has received only little attention in the past. The following chapter will review the available literature and relevant findings regarding the performance of people with LD as eyewitnesses and thereby highlight where additional knowledge is needed.

Chapter 3

The performance of people with LD as eyewitnesses

This chapter gives an overview of previous and current research regarding the performance of people with LD as eyewitnesses. The chapter firstly emphasises that although people with LD are likely to form a large witness population they might often be treated in a disadvantaged manner and excluded by the criminal justice system to give evidence, since research has shown that a significant amount of people regard individuals with LD as unreliable eyewitnesses (Stobbs & Kebbell, 2003). Thereafter, early and current experimental studies are described investigating eyewitness performance of people with LD and their findings are discussed. Finally, a review is provided of the special measures which are used by the legal system to ensure that witnesses and victims with LD are treated in an appropriate way when they have to give evidence at the police station or in court.

3.1 Introduction

In a case study reported by O' Hara (2001), which was mentioned briefly in Chapter 1, a woman with severe LD was referred to the local community team for people with LD for specific nursing support regarding a termination of pregnancy. It appeared that she had been sexually abused and that the pregnancy was a consequence of this abuse. These circumstances resulted in the involvement of the police. With the aid of sign language the woman was able to consistently identify the perpetrator, who was a family member. Furthermore, forensic evidence supported her allegations and the police took the case seriously. Unfortunately, despite the forensic evidence and the repeated identification of the perpetrator by the victim, the case did not reach court.

According to the law, the LD of the victim made her incapable of giving consent in court.

In another case reported by The Guardian (2008), several men were accused of raping and assaulting a woman with severe LD. However, there was no resulting prosecution, because again the victim was regarded as an unreliable witness by the Crown Office, which is responsible for the prosecution of crime in Scotland.

Although only two cases are mentioned here, they are useful case studies highlighting the importance of research to investigate the actual performance of people with LD as eyewitnesses. Moreover, they demonstrate how disadvantaged someone with LD might be treated by the criminal justice system when they become victims of or witnesses to crimes. This chapter gives an overview of previous and current psychological research examining the performance of individuals with LD as eyewitnesses. The following topics are addressed: attitudes toward witnesses with LD by lay people and the criminal justice system, the impact of different questioning formats on accounts given by people with LD, ways to improve eyewitness performance of people with LD and, finally, supporting measures available in legal proceedings for vulnerable witnesses.

3.2 Attitudes towards witnesses with LD

The two previously mentioned cases show clearly that the ability of individuals with LD to act as reliable and accurate eyewitnesses is often called into question by the criminal justice system. Several studies have investigated the perceived credibility of and attitudes towards witnesses with LD and found that people often possess rather negative opinions regarding the abilities of people with LD to act as reliable witnesses.

3.2.1 Police officers' attitudes towards individuals with LD

Bailey, Barr and Bunting (2001) evaluated police officers attitudes towards people with LD prior to and after awareness training. The awareness training was conducted by the Royal Ulster Constabulary. Training included role-plays and briefings from professionals. The Attitudes towards Mental Retardation and Eugenics (AMRE) questionnaire was used to measure attitudes towards people with LD before and after the training. The AMRE is a self-report rating scale that measures the extent of agreement and disagreement with 32 statements. It was established that the awareness training had a significant impact on AMRE scores. Police officers scored significantly higher on the AMRE after they had received awareness training. This indicates that they demonstrated a more positive attitude towards people with LD after the training. Although, these findings suggest that awareness training has a positive effect on police officers' attitudes regarding people with LD, Bailey et al. (2001) found some evidence that police officers possessed some discriminatory attitudes towards people with LD. This might have severe impacts on the progress and outcome of legal investigations. Since police officers are often the first people who will encounter information from witnesses with LD, they might also be the ones who make decisions about the credibility of the provided evidence and whether a case will be investigated further.

3.2.2 Jurors' attitudes towards witnesses with LD

In countries in which a jury system is still present, such as the United States and the United Kingdom, jurors' opinions about the reliability of a witness testimony might have serious implications for the outcome of criminal cases. Stobbs and Kebell (2003) examined jurors' perceptions of witnesses with mLD. During the study, mock-

jurors read transcripts of a mock-trial, including a testimony of an eyewitness with mLD or without LD. It was found that the eyewitness with mLD was perceived as significantly less credible, accurate, competent and 'good' in general than the witness without LD. However, responses indicated that mock-jurors also regarded the witness with mLD as the more honest and truthful witness.

Similar findings were obtained from a study conducted by Peled, Iarocci and Connolly (2004). The Credibility Comparisons Questionnaire (CCQ) was used to assess perceived credibility of child witnesses during a mock trial. Participants were presented with a testimony from either a 15-year-old child with mLD, a 15-year-old normally developing child, or a 10-year-old normally developing child. Mock-jurors rated the witness with mLD as less credible than the same age witness without LD. Furthermore, participants considered the witness with mLD as even less credible than the younger child witness without LD. This was a surprising finding, given that participants were informed previously that the cognitive level of a 15-year-old mLD child is similar to that of a 10-year-old normally developing child. Thus, the mere knowledge of an eyewitness having LD, seems to influence people's attitudes towards their credibility and reliability in a negative way.

3.2.3 Treatment of witnesses with LD in court by lawyers and judges

The research outlined so far has concentrated on attitudes of police officers and lay people towards witnesses with LD, but do these attitudes also manifest themselves in behaviours of professionals? Kebbell, Hatton and Johnson (2004) examined this research question in more detail by investigating which questions lawyers ask witnesses with LD in court and whether the same questioning strategies are applied with witnesses without LD. Real court transcripts were used as evaluation material from 16 rape, sexual assault and assault trials including witnesses with LD and 16

matched cases involving witnesses without LD. The trials were held in English courts between 1994 and 1999. Overall, it was found that lawyers asked witnesses with and without LD the same questions, including a large amount of closed questions, negatives, double negatives, multiple questions, and questions about numbers, names, times and dates, all of which are known to be very difficult to answer for people with LD (Plotnikoff & Woolfson, 2009), since they often show a high tendency of suggestibility and acquiescence (Clare & Gudjonsson, 1993). Consequently, such a questioning approach is likely to lead to unfavourable recall accuracy as well as perpetuating the perceived lower credibility of testimonies provided by witnesses with LD.

The fact that lawyers are likely to question witnesses with and without LD in a similar way seems to be controversial, but might it be that judges therefore more often intervene when a witness with LD is questioned? O'Kelly, Kebbell, Hatton and Johnson (2003) evaluated the extent and nature of the judicial interventions during court transcripts of 32 witnesses, of which 16 were witnesses with LD and 16 were witnesses without LD. Surprisingly, there were no significant differences in the amount of interventions made by the judge when the two groups were compared.

Thus, although a significant amount of people believe that individuals with LD are less credible and reliable witnesses than other members of the general population, there is no scientific evidence that the criminal justice system engages in different questioning strategies or additional guidance to support LD witnesses during court procedures. This is the case, even though research has repeatedly demonstrated that the utilised question format can have severe impacts on the accuracy of eyewitness accounts in general (Clifford & George, 1996; Fisher, Geiselman, & Raymond, 1987) and specifically on accounts provided by people with LD (Milne, 1999).

3.3 Research regarding eyewitness accounts of individuals with LD

Most experimental studies investigating the performance of people with LD as eyewitnesses involve children. Typically, participants are required to watch or engage in an event and are then subsequently interviewed about it to assess the quantity and quality of the reported event information.

The majority of studies use three distinct questioning formats: free recall, specific questions and leading questions. Free recall refers to open-ended questions, such as ‘What happened?’. Specific questions usually follow the free recall and aim to elicit more detailed responses from the participants and might be in the form of ‘What did the perpetrator wear?’. Leading questions are often used to assess the degree of suggestibility in individuals, since they include the required response. An example of a leading question would be ‘Did the perpetrator wear blue trousers?’.

3.3.1 Eyewitness accounts by children with LD

One of the first studies which examined the impact of different question formats on the recall memory performance of children with mLD was carried out by Dent (1986). The author investigated the impact of unprompted free recall, general questions and specific questions on the completeness and accuracy of the recalled information for a live staged event. It was found that children with mLD provided the least complete recalls during the free recall condition compared to the general and specific questions conditions. Similar amounts of event information were obtained during the general and specific questions conditions. With regard to the accuracy of the provided event information, it was found that children with mLD gave the most accurate accounts during the general questions condition and the least accurate accounts during the

specific questions condition. On the basis of these findings, Dent concluded that individuals with LD are not necessarily poorer eyewitnesses than individuals without LD, however, there is evidence that they respond optimally to different question formats. The findings of this study suggest that free recall questions should be regarded as the optimal interview technique for witnesses with LD.

A major short coming of this early study is that it did not include an appropriate control group, children without LD. Thus, although the study provided evidence that a free recall questioning format is the best approach to interview child witnesses with LD, it remains unclear whether this approach is in general the best interviewing method for children or whether it is particularly suitable for children with LD. A study which did include an appropriate control group was carried out by Henry and Gudjonsson (1999). They examined the ability of children with LD to recall event information and compared their performance with that of a chronological age (CA) matched control group and a mental age (MA) matched control group. By applying this methodological approach, Henry and Gudjonsson were able to further examine whether any arising difficulties in eyewitness memory performance in individuals with LD are due to developmental differences ('developmental approach') or the result of intrinsic differences caused by the LD itself (see *Handbook of Mental Retardation and Development* by Burack, Hodapp & Zigler (1998), for an in-depth discussion of the two theoretical models). The procedure utilised was similar to that used by Dent (1986). Participants viewed a live staged event and were questioned one day later about it during an interview. The interviews all started with a free recall phase. Thereafter, general questions, open-ended specific questions and closed yes/no questions followed in a hierarchical order. The closed questions included an equal amount of correct leading and misleading questions, thereby enabling an examination

of the amount of suggestibility in participants. The results showed that the amount of information obtained during the free recalls differed significantly between groups. The CA matched group recalled significantly more event information than the MA matched group. Children with LD did not differ significantly from either control group. No significant group differences were obtained for the other questioning formats. With regard to the accuracy of the obtained information no significant differences were found between the different groups during the free recall, the general questions, specific questions and the non-leading open-ended questions. Overall, the accuracy during the free recall was very high and only dropped to moderate levels during the more general question format. However, children with LD were more suggestible to misleading closed questions than their CA matched peers, but not than MA matched participants. Henry and Gudjonsson concluded that eyewitness memory of children with LD about a live staged event does not differ significantly in quality from that of children without LD. However, children with LD are more suggestible to closed misleading questions.

Michel, Gordon, Ornstein and Simpson (2000) used a more ecologically valid approach to investigate the ability of children with and without LD to remember information about an experienced event. Participants took part in a simulated health check, which can be regarded as a more interactive real-world experience, compared to watching a staged event. Furthermore, Michel et al. (2000) examined immediate as well as long-term memory performance of children with LD, by interviewing participants directly after the experienced event and 6 weeks later. Children in the control group were half matched on MA and half matched on CA. A standard interview protocol was used with questions organised in a hierarchical order, beginning with open-ended questions and progressing to more specific questions.

There were also questions included about features which did not occur, of which half were phrased in a neutral way and the other half in a suggestive manner. Overall, it was found that children in the CA matched group reported significantly more information than children with LD and the MA matched group. The recall performance of all participant groups declined after the six weeks delay. With regard to suggestibility to absent features presented during specific yes/no questions, it was found that LD and MA matched children were significantly more suggestible than the CA matched controls, which is in agreement with the findings obtained by Henry and Gudjonsson (1999). Furthermore, participants were more suggestible over time. No significant group differences were obtained for the accuracy of the recalled information. Consistent with the findings obtained by Dent (1986) and Henry and Gudjonsson (1999), accuracy dropped with the use of more specific question formats.

Thus, from the research findings obtained so far, it can be concluded that children with LD do at least show recall performance for an event which is in line with what would be expected of children with the same MA, but they do perform poorer compared to a CA matched control group. However, slightly different results were obtained by Agnew and Powell (2004). An interactive stimulus event was used to investigate the impact of different question formats on recalls of children with LD. The performance of children with LD was compared with that of a CA matched and a MA matched control group. Since the LD group was quite large and included children with mild as well as moderate LD, it was further possible to examine the influence of the severity of the LD on the eyewitness performance of LD participants. After interactively participating in a 30 minute magic show, children individually attended two interviews. The purpose of the first interview was to suggest true and false information. The second interview was designed to test memory performance of the

children. The interview consisted of a free narrative phase followed by a series of open-ended questions. For each item that was not recalled during the free recall phase the interviewer asked one specific cued-recall question. If the required information was nevertheless not obtained, in response to the specific cued-recall question, the child was provided with a forced-choice question. Agnew and Powell (2004) found that children with LD reported significantly fewer items of correct event information than did children of the two control groups. Moreover, children with LD needed greater specificity of questioning than children in the CA and MA matched control group. No differences were obtained regarding the completeness of the reported information and the severity of the LD. For the free narrative phase of the interviews, no significant group differences were obtained in the accuracy of provided information, which is in agreement with the findings obtained during the earlier mentioned studies. For the specific cued-recall questions and the forced-choice questions, a significant effect for group was obtained; participants with LD recalled a smaller proportion of accurate responses compared to both the CA and the MA matched control groups.

These findings seem to contradict the ones obtained by Henry and Gudjonsson (1999) and Michel et al. (2000), since they have found that children with LD show at least eyewitness abilities which are appropriate for their mental age. Another contradicting finding was revealed with regard to children's suggestibility. In the Agnew and Powell (2004) study, children with LD were not found to be more suggestible than control participants; in fact, they were significantly less likely to repeat false suggestions introduced by the interviewer than children of the two control groups. However, they did provide a larger proportion of external intrusion errors, i.e. reporting entirely new false items that were non existent and not suggested by the

experimenter. The recall by LD children also included significantly less specific event information than recall provided by the two control groups. No significant differences were obtained in completeness or accuracy of the recalled event information during any of the questioning formats and the degree of the LD (mild vs. moderate). According to Agnew and Powell (2004), the obtained findings indicate that children with LD are able to provide accurate information about an experienced event, however, they were providing less complete as well as less accurate accounts of the event and they required more specific questioning than children without LD. These findings do not support those obtained by Henry and Gudjonsson (1999) and Michel et al. (2000). However, it should be noted that during these two studies only children with mLD participated, whereas Agnew and Powell (2004) included a much more diverse participant group, including children with mild as well as moderate LD. Although Agnew and Powell (2004) did not obtain any significant differences in eyewitness performance between the mild and moderate LD participant groups, it can be argued that the inclusion of the moderate LD group led to a decrease of the overall LD group performance and therefore they differed significantly from the CA as well as the MA matched control groups.

Taking all of the above into consideration, it is important to note that all studies have shown that children with LD can provide accurate and potentially useful information about a perceived event, which is at odds with the generally held view of lay people and the criminal justice system, that individuals with LD are as such unreliable and inaccurate eyewitnesses.

3.3.2 Eyewitness accounts by adults with LD

All the studies cited so far have investigated the ability of children with LD as eyewitnesses, but do the same rules apply for adult witnesses with LD? Perlman,

Ericson, Esses and Isaacs (1994) investigated the impact of different question formats on observed-event accounts provided by adults with mLD and without LD. Participants viewed a 7-minute video clip of a crime and were subsequently questioned with five different question formats: free recall, general questions, short answer questions, specific questions and statement questions. The findings revealed that during the less structured recall formats, such as the free recall questioning and the general questioning format, individuals with mLD provided significantly less information about the witnessed event than participants without LD. However, the information provided by the mLD group was not significantly less accurate than that provided by the control group. The short-answer questions elicited significantly less accurate responses from both participant groups compared to the more general recall formats, but particularly from participants with mLD. With regard to correct leading questions, both participant groups performed equally. However, participants with mLD had significantly more difficulties with questions including misleading information. On the basis of these findings, Perlman et al. (1994) concluded that a combination of free recall questions and specific questions would elicit the most complete and accurate accounts about a to-be-remembered event from individuals with mLD. Leading questions, specifically those including false information should be avoided when questioning witnesses with LD.

In summary, people with LD can give reasonably accurate accounts of perceived events, when an open-ended question format is used. However, they also tend to provide only sparse amount of information. A similar pattern of findings was obtained for adults as with children. Consequently, interviewers might feel forced to ask more specific, closed-ended questions to obtain all relevant information. Unfortunately,

these question formats are also the ones which elicit the least accurate information from individuals with LD, which might be partially due to their heightened level of suggestibility. Although research has repeatedly demonstrated that open-ended recall questions should be used and specific and suggestive questions should be avoided, it is surprising and worrying to see that in practice the police are still relying frequently on specific questions rather than using more open-ended ones (Cederborg & Lamb, 2008).

3.4 Ways to improve eyewitness accounts of LD individuals

Research has established that people with LD can provide accurate accounts of a witnessed event, if questioned in an appropriate way. However, their accounts are often incomplete. Therefore, it is important to examine whether there are ways to improve the quantity and quality of their eyewitness accounts.

3.4.1 The Cognitive Interview

An interviewing technique particularly suitable for witnesses with LD is the Cognitive Interview (CI). The CI was developed by Ed Geiselman (University of California, Los Angeles) and Ron Fisher (Florida International University) in 1984. It includes several memory retrieval techniques also known as mnemonics, which aim to increase the quantity and quality of remembered information from eyewitnesses (Geiselman et al., 1984). The mnemonics are based on fundamental theoretical principles regarding memory organization, storage and retrieval, such as Tulving and Thomson's Encoding Specificity Hypothesis (1973) and Bower's (1967) multiple-component memory trace theory. According to Tulving and Thomson (1973), successful retrieval of information is most likely when the context and the cues present at retrieval match those present at the initial encoding. Therefore, the reinstatement of the initial encoding context

should lead to an increase of the accessibility of the stored information. Bower's multiple-component memory trace theory (1967) proposes that our memory is a network of associations rather than an accumulation of single, unconnected incidents. Consequently there are multiple ways to access or cue one specific memory. Initially, the CI consisted of four general retrieval techniques: the mental context reinstatement of the personal and physical context, present at the time the witness experienced the event; the change perspective technique, during which the witness is encouraged to place themselves in the shoes of the victim or another witness; the report everything instruction, during which the witness is encouraged to report everything he/she can remember including partial information; and finally the reverse order technique, during which the witness is asked to make several retrieval attempts from different starting points in time (Memon & Koehnken, 1992). In addition to the cognitive retrieval techniques, the CI also includes more social/communication techniques, such as rapport building, transferring of control and report everything instruction, which aim to facilitate communication between the interviewer and the interviewee. The mnemonics and the social/communication techniques play an important role during the interview and complement one another (Memon, 1997).

The CI has been empirically tested during numerous studies (see Koehnken, Milne, Memon & Bull, 1999 for a detailed meta-analysis of the usefulness of the CI or Griffiths & Milne, 2010 for a recent review chapter) and across various populations, including children (Milne & Bull, 2003), people with LD (Milne, Claire & Bull, 1999) and the elderly (Wright & Holliday, 2007) and its effectiveness has been well established. The original CI was further refined during the subsequent years and resulted eventually in the Enhanced Cognitive Interview (ECI). The ECI puts even more weight on the social and communication aspects of the interview, such as

timing of questions, appropriate interviewing environment and meaningful structure of questions (see Dando & Milne, 2009 for a detailed review on the ECI).

So far, four studies have been conducted that investigated the effectiveness of the CI with individuals with LD. The earliest study was carried out by Brown and Geiselman in 1990. Participants, comprising adults with and without LD and children, were presented with a video and after a two-day time delay they were questioned either with the CI or a standard interview (SI). The number of correctly recalled items, incorrect items and confabulations were recorded. The results showed that the CI was very successful in gathering information from participants with LD. Although adults with LD recalled in general less information than the two other participant groups, regardless of the type of interview, the CI increased the number of correct items recalled, without increasing the number of incorrect items in comparison to the SI. However, the CI did produce more confabulations in participants with LD than the SI. Overall, the findings of that study appear to be very promising, since the use of the CI enhanced the recall of information by 32% compared to the SI in individuals with LD. It should be noted that individuals with LD benefited even more from the CI than children, who showed only an increase of 21% of correct information. Adults with LD performed similar to adults without LD with regard to the accurate amount of obtained information during the CI.

The first study which investigated the effectiveness of the CI for the use with children with mLD was carried out by Milne and Bull (1996). Participants watched a video of a magic show and were interviewed one day later with either the CI or the SI. It was found that children interviewed with the CI recalled significantly more correct information, without significantly increasing the amount of incorrect information. Furthermore, it was examined whether the CI might reduce the effect of suggestive

questioning in children with mLD. Therefore, a questionnaire was administered either before or after the interview, comprising leading, misleading and non-leading questions. The findings showed that after the CI, children were more likely to resist misleading and leading questions, however, these effects did not reach significance. On the basis of these findings the authors concluded that the CI has the potential to be a useful interviewing tool with children with mLD in an legal setting. The findings of this study are in agreement with those obtained by Brown and Geiselman (1990) with adults with LD.

A further study investigating the usefulness of the CI with adults with mLD was conducted by Milne et al. (1999). Adults with mLD and without LD took part during this study. Participants viewed a three minute video depicting a car accident. One day later participants were either interviewed with the CI or the SI. It was found that the CI produced significantly more correct recall than the SI in both the mLD as well as the non-LD participants; approximately 35% more correct information was elicited from the LD group and 20% from the control group. However, for the participants with mLD, the CI also elicited significantly more confabulations, which is consistent with the findings revealed by Brown and Geiselman (1990).

Similar to Milne and Bull (1996), Robinson and McGuire (2006) were interested in the effect of the CI on suggestibility. Children with and without mLD took part in the study. The study was divided in two parts. During the first part, the Gudjonsson Suggestibility Scale 2 (GSS2) was administered to assess the level of suggestibility in children. In the second part, children watched a short video clip and were subsequently interviewed with either the CI or the SI. Subsequently, participants were asked to answer several questions about the observed clip, including non-leading as well as misleading questions. A week later, participants were again questioned

about the clip, once more the questions included non-leading and misleading ones. It was found that children with mLD scored higher on all measures of suggestibility on the GSS2, although not significant, these trends indicate that children with mLD display a higher level of suggestibility than children without LD. The CI elicited significantly more correct information from the children than the SI. However, children also reported significantly more incorrect information and more confabulations when interviewed with the CI. No supportive evidence was found for the assumption that the CI would reduce the amount of suggestibility in children, particularly in those with mLD. Children interviewed with the SI were misled by 68.4% of the misleading questions and children interviewed with the CI were misled by 67.5% of the misleading questions. The findings are contradictory to the ones obtained by Milne and Bull (1996), who found that the CI seems to reduce the amount of suggestibility in children with LD. However, it should be noted that Robinson and McGuire (2006) applied a much longer delay between the interviews and the subsequent second questioning phase. Future research into the possible interaction of delay and suggestibility is needed.

In summary, the findings of all these studies are in agreement with those investigating the impact of different question formats on accounts by individuals with LD. Accounts were typically less complete than those provided by individuals without LD, but not necessarily less accurate. People with LD showed similar benefits from the CI as people without LD, namely a significant increase of reported information. This finding confirms that the CI might be a suitable interview tool for witnesses with LD. However, it should also be noted that there is still room for further improving techniques to interview witnesses with LD, as several studies have shown that the CI increased the amount of confabulations and incorrect information reported.

Furthermore, it does not seem to successfully prevent from the impact of suggestive questioning.

3.4.2 Repeated questioning of individuals with LD

On occasions, it might be useful and/or necessary to re-interview witnesses so that they can elaborate on the information they have already provided during the first interview and provide information about topics which have not been discussed yet, although some studies have found that people become more inaccurate during repeated questioning (see Moston, 1990 for a review), especially children (Leichtman & Ceci, 1995; Poole & White, 1991). This is most frequently the case when misleading and closed-ended questions are used. As discussed earlier during this chapter, research has shown that these kinds of questions are in general viewed as inappropriate and should be avoided. When open-ended questions are used, repeated questioning appears to have no detrimental effect on the accuracy of individuals' accounts (Poole & White, 1991). Some studies even found a beneficial effect of repeated interviewing when an open-ended question format was employed (La Rooy, Pipe & Murray, 2005; Memon & Vartoukian, 1996), thus an increase in correctly reported event information.

Only a few studies have looked at the effect of repeated interviewing on accounts provided by individuals with LD. Henry and Gudjonsson (2003) interviewed children with mild and moderate LD about a live staged event one day or two weeks later. The interviews were organized in a hierarchical order, starting with a free recall and followed by increasingly specific questions. Closed ended yes/no questions at the end included correctly leading as well as misleading ones. The study included CA and MA matched control groups. All three participant groups showed an increase in the amount of reported correct information during the free recall during the repeated

interview compared to the initial interview. However, LD children changed their answers significantly more often during the repeated interview than did MA and CA matched control children. This might have severe consequences in a legal context, as it may lower the perceived credibility of the witness.

Cederborg, La Rooy and Lamb (2008) examined 20 interviews with alleged child victims with mild or moderate LD. The interviews were obtained from a larger project from the Swedish police including in total 69 criminal cases. They were selected on the basis that they featured examples of child witnesses having been interviewed twice. During the analysis of data, the quality of the first and second interview was assessed. The study revealed that about 80% of the information produced during the repeated interview was completely new and about forensically relevant topics. Only a small amount of information obtained during the repeated interview contradicted information initially reported.

Thus, it appears that repeated interviews can be valuable in a legal setting with individuals who have LD, when they are conducted in an appropriate way. The above mentioned research indicates that providing witnesses with LD with a second chance to tell about their experiences can generate new and potentially useful information to the police. However, it should be kept in mind that the quality of the information obtained during the repeated interview, to a certain extent, depends on the questioning skills of the interviewer and the type of questions asked.

3.5 Special legal measures for witnesses with LD

In recent years, the interests of vulnerable witnesses in legal proceedings, including witnesses with LD, have received more and more recognition by the criminal justice system in the UK. This might be partially due to the increase in scientific research

regarding the performance and special needs of vulnerable witnesses. As a result, the Home Office published the document *Speaking up for Justice* in 1998, which was composed by the interdepartmental Working Group on the Treatment of Vulnerable and Intimidated Witnesses. It includes multiple recommendations for the criminal justice system on how to treat vulnerable and intimidated witnesses in a fair manner and how to assist them optimally to give best evidence during legal proceedings. Several of the recommendations were subsequently implemented in *The Youth Justice & Criminal Evidence Act (1999)*, which got introduced in England and Wales (Ellison, 2001). Most of the special measures included in the Act stem from the domain of child witnesses and are now also available for adult vulnerable witnesses, such as witnesses with LD (Cooke & Davies, 2001). The special measures included in the Act apply predominantly to court proceedings and include the following:

- S.23 Screens. Witnesses can be provided with a screen to prevent them from being confronted with the accused.
- S.24 Live link. The witness can give evidence during the trial via a close-circuit television link to the court room.
- S.25 Exclusion from court. In cases of intimidation or sexual assault, the public and the press can be excluded from the trial.
- S.26 Removal of wigs and gowns. Judges and barristers might be asked to remove their wigs and gowns on behalf of the witness.
- S.27 Video evidence-in-chief. The witness might be permitted to give the evidence-in-chief on videotape prior to the court case.
- S.28 Video cross-examination. When the witness gave the evidence-in-chief on videotape prior to the court trial, they might be also permitted to be cross-examined and re-examined on videotape prior to the court trial.

- S.29 Use of intermediary. An intermediary might be appointed by the court to facilitate communication between the witness and the court during the trial.
- S.30 Aids to communication. Aids to communication might be available to the witness, such as sign and symbol communication boards or electrical equipment (Cooke & Davies, 2001).

In Scotland, The Vulnerable Witness (Scotland) Act 2004 was implemented recently and it shares several similarities with The Youth Justice & Criminal Evidence Act 1999. The Act includes three major legislative changes: first, it includes a much wider definition of the term “vulnerable witness”, comprising now anyone where there is a considerable risk that the quality of evidence may be diminished by reason of fear or distress with regard to giving evidence; second, child witnesses now have an automatic entitlement of the use of special measures; third, the court is permitted to use special measures also in civil cases; and finally, the Act abolished the pre-testimony competence test for witnesses in criminal and civil proceedings (Sharp & Ross, 2008).

In addition to the recommendations applicable during the trial, several recommendations in the Speaking up for Justice report (1998) were specifically addressed to police officers to assist them in the fair treatment of vulnerable witnesses during the investigation stage. The recommendations for police officers include amongst others:

- The police should aim to identify vulnerable witnesses as early as possible during the investigation process.
- Police services should identify individuals who have received special training in dealing with vulnerable witnesses to assist in the identification of those witnesses.

- A series of prompts should be developed by the Association of Chief Police Officers (ACPO) in collaboration with the Department of Health and the Disability Policy Division to aid the overall assessment of an individual witness' needs. The resultant prompts can be found in the *Vulnerable Witnesses: A Police Service Guide* (ACPO & Home Office, 2002).
- The police should seek advice from those who know the vulnerable witness best on how to best communicate with him/her.
- The police should be responsible that a supporter is present during the interview of a vulnerable witness, e.g. "appropriate adult".
- The vulnerable witness should have a say in which pre-trial and trial measures are employed.
- The police should have early meetings with the Crown Prosecution Service (CPS) regarding special measures needed to assist the vulnerable witness before and during the trial.

Other relevant documents particularly composed to assist police officers are *Vulnerable Witnesses: A Police Service Guide* (2002) and *Achieving Best Evidence in Criminal Proceedings: Guidance on Interviewing Victims and Witnesses, and Using Special Measures* (2007). The former is a comprehensive guide for police officers on how to identify vulnerable witnesses at the investigation stage and the latter provides specific advice on how to best plan and conduct interrogative interviews with different vulnerable witness groups, such as intimidated-, reluctant-, hostile- and defence witnesses. It should be noted that all these documents are merely advisory and do not constitute a legally enforceable code of conduct. It therefore remains questionable whether all UK police officers make reference to these guidelines and if

so in what way and during which situations. For the future, it would be desirable to see that not only special measures for vulnerable witnesses during court procedures become statutory, but also measures applicable during the investigation and interrogation process.

3.6 Discussion

Although the majority of people appear to believe that individuals with LD are more unreliable witnesses than individuals without LD, research which has investigated eyewitness accounts of people with LD has repeatedly demonstrated that they can give very accurate reports of an experienced event when the question format takes their LD into consideration. Nevertheless, caution should be taken when interviewing people with LD because of their high susceptibility to suggestions and their increased tendency to acquiescence. Therefore, specific and suggestive questions should be avoided and a more open-ended questioning style should be applied when interviewing witnesses with LD. An appropriate interviewing technique for people with LD appears to be the CI, since it increases the amount of correct reported event information without necessarily increasing the amount of incorrect information. The CI, or elements of it, is also applied during several experimental studies comprised in this thesis. If necessary, re-interviewing witnesses with LD can be valuable as well, since research has shown that during repeated interviews new and forensically relevant information can be obtained. However, the quality of the information gained during the repeated interview seems to depend not merely on the recall abilities of the witness but partially as well on the questioning skills of the interviewer.

Over the last few years, a lot of research has been conducted regarding the eyewitness performance of people with LD, and many improvements have been made

in UK legislations, including the introduction of special measures to support LD victims and witnesses. There remain several aspects which need to receive further examination. Up until now, the majority of studies which have investigated the eyewitness performance of individuals with LD, have focused on their ability to recall information about a-to-be remembered event. However, an important task of a witness who has observed a crime is to describe the face of the perpetrator to the police. The ability to recognise and describe a face might be crucial, for example during the accurate construction of a facial composite image. A comprehensive search of the literature, using search engines, such as PsycInfo, Web of Knowledge, ZETOC, etc., including keywords, such as “learning disability”, “intellectual disability”, “witnesses”, “face recognition”, “facial composites”, etc., revealed no published research which has investigated the ability of people with LD to use facial composite systems and to generate accurate and reliable composite images. A facial composite image can facilitate the criminal investigation and might aid to the successful solution of a crime. People with LD seem to be even more likely, due to their increased susceptibility to victimisation, to encounter situations in which they are required to describe a perpetrators’ face to the police; it is therefore surprising to see that previous research has somewhat neglected this topic. The current PhD project therefore aims to systematically investigate basic face recognition and description abilities of individuals with mLD and their performance with facial composite systems which are currently used by the UK police. By doing so, the project will add to the existing base of knowledge regarding the eyewitness performance of individuals with LD.

Chapter 4

Review of applied research in face recognition and recall

Before proceeding to a more practical issue of face recognition and description, namely the construction of facial composites (which is described and discussed in more detail in Chapter 5) it is important to consider previous and recent research findings and influential theoretical frameworks regarding face recognition and recall. This chapter gives a brief overview of research and theories focusing on forensic aspects of face memory. First, situational factors influencing face recognition are discussed, such as the effects of viewpoint, lighting and context. Second, factors inherent to faces themselves are addressed, such as distinctiveness and featural versus configural face processing. Finally, research regarding face recall is summarised, emphasising relevant forensic aspects such as feature saliency and the verbal overshadowing effect.

4.1 Introduction

Everyday experience attests that humans have the capacity to perceive the unique identity of a virtually unlimited number of different faces. It is not too surprising then that much of the research on face perception has focused on this ability to discriminate and recognise individuals. The 1970's saw an expansion of this type of research, particularly of research regarding face recognition (Ellis, 1975). Since then, thousands of studies have been conducted investigating not only human face recognition but also face recall (Ellis & Shepherd, 1992). The current chapter, however, focuses particularly on applied research. Thus, it specifically addresses

work targeting forensic aspects of face memory, such as eyewitness identifications and the accuracy and completeness of witnesses' facial descriptions. Given the extensive amount of relevant literature, the aim of this chapter is to give a comprehensive but concise overview of influential theoretical frameworks as well as previous and current research findings.

In an eyewitness situation, such as during the construction of facial composites, witnesses are often required to engage in multiple tasks involving face recognition, recall and description (Pike, Kemp & Brace, 2000). Errors in suspect identifications or descriptions can have severe consequences. The Innocence Project (2010)¹ states that the single greatest cause for wrongful convictions worldwide are eyewitness misidentifications. Such mistakes have been found to be responsible for around 75% of convictions which were later exonerated by DNA evidence. A well-known UK case, involving an eyewitness misidentification, was the Jill Dando murder in 1999. The prime suspect was Barry George, a local man with a history of criminal records. Barry was convicted of the crime in 2001 and the eyewitness evidence was deemed to be an important piece of evidence presented in court. However, it turned out that this testimony might have been fallible due to prior discussion between co-witnesses. In 2007 the conviction was quashed and in 2008 Barry was acquitted due to insufficient evidence (Davies & Griffiths, 2008; Wright, Memon, Skagerberg, & Gabbert, 2009).

In contrast, only a few criminal cases are known during which wrongful face recall played a role (Shepherd & Ellis, 1996). For instance, erroneous face recall may have slowed down dramatically the criminal investigation during the 'Baton Rouge Serial Killer' case in 2002. Derrick Todd Lee murdered seven women and was

¹ Retrieved from The innocence project <http://www.innocenceproject.org/understand/Eyewitness-Misidentification.php> on the 3rd of August 2010

convicted in 2002 on the basis of DNA evidence. During the investigation several composites had been created some of which are now known to be very poor likenesses of him. In hindsight, critics speculate that some of these poor composites misled the police and potential witnesses and contributed to a slowdown in the criminal investigation². These cases highlight the relevance of applied face recognition and recall research and its important contribution to the criminal justice system.

A review of the available literature reveals that most research has focused on face recognition rather than facial recall (Shepherd & Ellis, 1996). One reason for this might be that in everyday life people engage much more frequently in tasks that call on face recognition rather than being required to verbally describe a face (Ellis & Shepherd, 1992). Notwithstanding this observation, face recall should not be considered less important. There are obvious circumstances where its accuracy and completeness may have crucial impacts on the outcome of criminal cases. Both face recognition and recall draw upon our memory for faces; however, they both have their differences. As a result one should urge caution when making generalisations from the findings of one task to the other (Shepherd & Ellis, 1996). Consequently, this chapter first describes and discusses research addressing forensic aspects of face recognition and thereafter concentrates on research examining face recall. It should be noted that many factors relating to face recognition and recall accuracy have been investigated, but this chapter reviews only aspects relevant to the research conducted in this thesis.

² Retrieved from trutv crime library 'Derrick Todd Lee-The Baton Rouge Killer' by R. Bell http://www.trutv.com/library/crime/serial_killers/predators/baton_rouge/4.html on the 4th of August 2010

4.2 Face recognition

Factors that influence face recognition can be divided into two major categories: situational factors, such as viewpoint and lighting, and factors which are inherent to faces themselves, such as configurations and distinctiveness (Hancock, Bruce & Burton, 2000). The impact of these factors might be quite different for familiar and unfamiliar faces, since research suggests that different processes are involved in the recognition of familiar and unfamiliar faces (Johnston & Edmonds, 2009).

4.2.1 Familiar versus unfamiliar face recognition

The experimental studies comprised in this thesis use unfamiliar faces as stimulus material. They can be regarded as more ecologically valid, given that in a real eyewitness situation the perpetrator may often be unknown to the witness or victim, particularly when it comes to the construction of a facial composite. Nevertheless, it is important for researchers in the face recognition and recall domain to be aware of differences in processing of familiar and unfamiliar faces and the different impacts factors may have on these two categories of faces. This is especially important during the decisions of potential stimulus material and later during the interpretation of the results.

Evidence that familiar and unfamiliar faces are processed differently comes mainly from neuropsychological studies with patients who have face recognition impairments (Bruce, Burton & Hancock, 2007). The most well-known impairment is prosopagnosia, which is characterised by the inability to recognise familiar faces, such as the faces of relatives, acquaintances or famous people (Johnston & Edmonds, 2009). Interestingly, a number of studies have revealed dissociations in prosopagnostic patients between their performance on tasks involving the recognition

of familiar and unfamiliar faces. Benton and Van Allen (1972), for instance, reported a case of a patient with severe prosopagnosia who was still able to perform in the normal range when asked to match unfamiliar faces. Similar findings were reported by Malone, Morris, Kay and Levin (1982). One of their patients recovered from his inability to recognise familiar faces but remained incapable of matching unfamiliar faces, whereas another showed a persistent inability to recognise familiar faces but a gradual improvement in his ability to match unfamiliar faces up to a level in the normal range.

Comparable double dissociations have also been reported by Young, Newcombe, de Haan, Small, and Hay (1993) who examined the abilities of ex-servicemen with unilateral posterior brain lesions to perform tasks involving familiar face recognition, unfamiliar face matching and facial expression analysis. The authors found evidence for selective impairments of each of the three abilities in several of their patients. For example, one serviceman demonstrated impaired familiar face recognition only, whereas another one was incapable of accurately matching unfamiliar faces despite performing without restrictions on all other tasks.

These clinical observations clearly suggest that there is a distinction between familiar and unfamiliar face processing and they appear to be to some extent independent from one another (see Johnston & Edmonds, 2009 for a recent review on familiar and unfamiliar face recognition). The first theoretical frame work which aimed to explain the obtained dissociations between familiar and unfamiliar face recognition was developed by Bruce and Young in 1986 (Bruce et al. 2007). According to their functional model, face information is processed in a sequential order by different functional face processing components. Put differently, when we see a face we can derive several distinct types of information from this face. Bruce

and Young (1986) refer to them as pictorial, structural, visually derived semantic, identity-specific semantic, name, expression and facial speech information. They assume that when we see a face, a pictorial code is generated, which can be understood as something like a visual description. Necessarily, this pictorial code also contains information about lighting, orientation and expression. Matches at the level of the pictorial code can be used during old/new face recognition tasks in which the exact same facial stimuli are used at the learning and test phase. These types of tasks are also used during some of the experimental studies in the present thesis. Structural codes on the other hand capture more abstract information about the face, which makes it possible to recognise a face even when changes were made to its orientation or expression. Structural codes are necessary for recognising familiar faces.

Interestingly, people can draw some inferences from unfamiliar faces, such as judging their age and sex, using only visually derived semantic codes. The opposite of visually derived semantic codes are identity-specific semantic codes, which include information about a persons' occupation, friends and interests, etc. The identity-specific semantic codes are responsible for the 'feeling of knowing someone' and are only drawn upon in familiar face processing. After having recognised a person as someone familiar, it is often possible to generate the name of this person via a specific name code. Finally, expression and speech codes make it possible for us to interpret facial expressions and speech for both familiar and unfamiliar faces. However, these codes do not constitute important components in recognising faces.

Thus, Bruce and Young (1986) suggest that faces possess different types of information, such as pictorial and structural information and information about expression and speech. Several independent components are responsible for processing this information. The recognition of familiar faces can be understood in

terms of getting sequential access to the different codes. Figure 4.1 displays the functional model of face recognition.

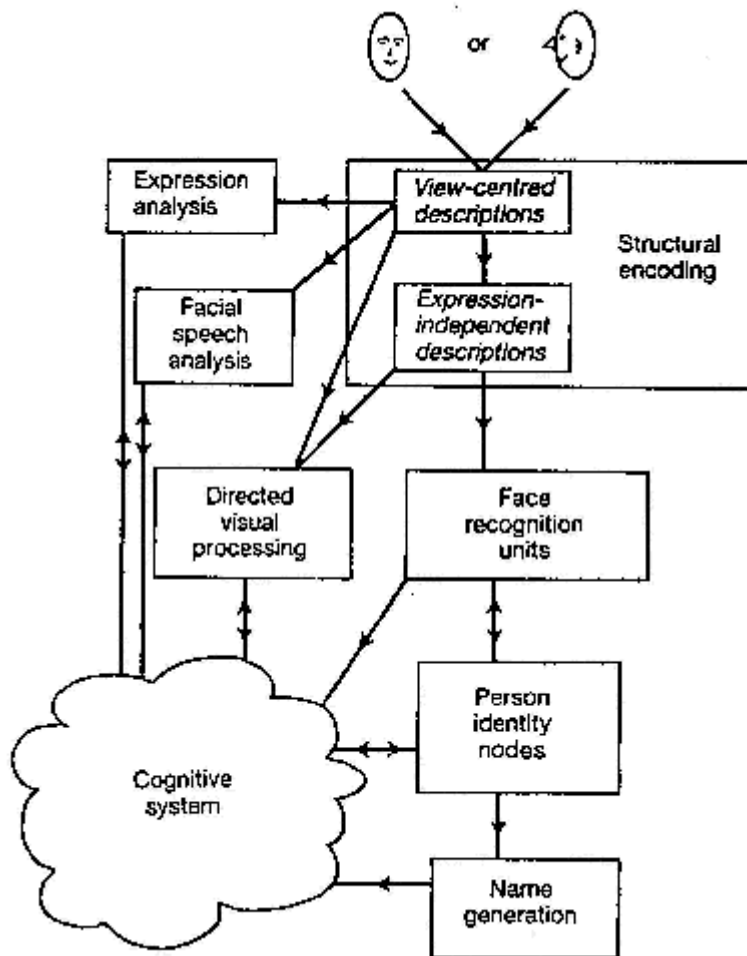


Figure 4.1. Bruce and Young's (1986) functional model for human face recognition (reproduced from Bruce, V., & Young, A. (1986). *British Journal of Psychology*, 77, 305-327).

The model provided one of the first theoretical frameworks for human face recognition and produced numerous research questions that formed the basis for much of the experimental work carried out in the 1980's and 90's. Furthermore, it explained the perceived dissociations between familiar and unfamiliar face recognition in prosopagnostic patients by proposing that different independent functional components are involved in the recognition of unfamiliar and familiar faces. Moreover, the model accounts for different impacts situational and face inherent

factors have on familiar and unfamiliar face recognition. Consider for example the impact of orientation and expression on unfamiliar and familiar face recognition. As mentioned earlier, unfamiliar faces are recognised when there is a match between the encoded representation and the pictorial code, which was formed during the initial encounter. Since the pictorial code can be regarded as a static image, slight changes to the orientation and/or expression can lead to a mismatch between the perceived image and the pictorial code, although it is still the same face. Familiar faces on the other hand are recognised via structural codes which contain additional abstract information, such as character attributes and information about a person's occupation etc. This information can be derived even if changes in orientation or expression have been made and therefore a face can be still perceived as familiar even if it is not an identical copy of the image initially encountered.

If there are differences in the way familiar and unfamiliar faces are processed one should be mindful of this when planning experimental procedures and interpreting results of studies that draw upon tasks where stimuli can be familiar or unfamiliar. With regard to the experimental studies in this thesis, during which unfamiliar faces were used as stimuli, it was critical to ensure that presented pictures of target faces were identical during the learning and test phases, to control for any confounding factors. Some of these influencing factors are reviewed during the next paragraphs.

4.2.2 Situational factors influencing face recognition

Situational factors are variables such as exposure time, delay, viewpoint, movement and degradation. They are associated with the surroundings and circumstances in and under which a face is perceived and can influence subsequent memory for this face. Whereas it is nearly impossible to control situational factors in the real world,

laboratory studies enable scientists to manipulate their occurrence and to investigate their distinct or combined impact on face recognition and recall (Sporer, Malpass and Koehnken, 1996). As mentioned earlier in the Introduction, this Chapter covers only research topics which are relevant to the current thesis. Therefore, several factors, such as movement and degradation will be omitted. This does however not mean that they should be considered as less important.

Exposure duration and delay

Exposure duration and delay can both be regarded as situational factors (Narby, Cutler & Penrod, 1996). Exposure duration can refer either to the amount of time the target is presented at the initial presentation or at the subsequent test. Delay refers to the duration of time between the first presentation of the target and the subsequent test phase.

A number of experiments have reported detrimental effects of a decrease in exposure time on unfamiliar face recognition (Krouse, 1981; Laughery, Alexander & Lane, 1971; Reynolds & Pezdek, 1992). For instance, Laughery et al. (1971) presented participants with pictures of faces for 10 seconds or 32 seconds, before they engaged in a subsequent face identification task. The results revealed that a longer target exposure was associated with better identification performance. Even more persuasive results were obtained by Reynolds and Pezdek (1992), who used composites, created with Identi-Kit, to investigate the impact of exposure duration on subsequent matching accuracy. During the matching task, participants were presented with 20 composites. The exposure duration was either three or 20 seconds. After a two minute delay the test phase commenced during which 40 composites were presented, of which 20 were old and 20 new (the new composites differed in one feature from the old ones). Participants were required to indicate whether the

presented composites at test were identical or different to the ones viewed at presentation. Reynolds and Pedzek (1992) found that the false alarm rate was significantly lower for the long exposure condition compared to the short exposure one. In combination these studies indicate that people who have more time to look at a face will generally recognise it better afterwards.

Not only can a short exposure towards a face have a disadvantageous effect on subsequent recognition accuracy, but so too can the delay time between the initial presentation and the attempt to recognise a face. In a study conducted by Walker-Smith (1978) the impact of delay between presentation and test and of exposure duration of the target face at test was investigated with the use of PhotoFIT composites. A similar procedure was used as by Reynolds and Pedzek (1992). Participants engaged in a matching task, during which pairs of identical or different composites were displayed consecutively. Participants had to make same/different judgments and the error rate of these decisions was calculated. During the short exposure condition, the test face remained visible for 65 milliseconds and for 1 second during the long exposure condition. Moreover, the experiment included a short (one second) and long delay condition (20 seconds). The analysis of the data revealed significant main effects for exposure at test and delay such that participants made considerably fewer recognition errors during the short delay and long exposure conditions. Thus, both exposure duration at presentation and test as well as delay between presentation and test can influence how accurate people remember and subsequently recognise unfamiliar faces.

On the other hand, long periods of delay appear to have only minimal impact on recognition accuracy for familiar faces. Bahrick, Bahrick and Wittlinger (1975) used a cross-sectional approach to investigate the impact of delay on peoples'

memory for familiar faces. Participants were required to recognise pictures of their graduating colleagues after delays ranging from 3.26 months to 47.56 years. Portraits were obtained from year books and presented to participants in the form of a recognition test. At test stimuli pairs were presented, displaying portraits of graduate colleagues accompanied by unfamiliar foils and participants were required to identify the faces of their colleagues. Bahrick et al. (1975) revealed that participants were able to accurately recognise 90% of their graduate colleagues even after a time delay of 35 years. So therefore it is obvious that variables, such as delay, exposure time, and familiarity not only have a distinct effect of their own but they can also have a combined effect as evidenced by the presence of significant interactions in factorial experiments.

Viewpoint and context

Several studies suggest that face recognition is strongly viewpoint and context dependent and that generalisation from one viewpoint or context to another can be experienced as very difficult (Bruce, 1982; Hill, Schyns & Akamatsu, 1997; Davies & Milne, 1982; Ewbank & Andrews, 2008; see also Zhao, Chellappa, Phillips & Rosenfeld, 2003 for a review). As with exposure and delay, this is particularly true for unfamiliar faces, whereas familiar faces appear to be less dependent on changes in position and context. Bruce (1982), for instance, changed the viewpoint (frontal vs. profile) of familiar and unfamiliar faces during an old/new recognition paradigm. She revealed that for unfamiliar faces the change in viewpoint had a significant impact on recognition accuracy and decision latency. Thus, the hit rate during test dropped while the time it took perceivers to make a decision increased. In contrast, altering the viewpoint of familiar faces between presentation and test had no measurable effect. Similarly, Hill et al. (1997) revealed significant interactions between learning view

and test view on recognition accuracy for three-dimensional face models. For example, for learned full-face views, an inverted U shape recognition pattern emerged, with performance becoming poorer with increasing angle of rotation from the learned view.

Not only can changes in viewpoint have a considerable impact on the accuracy and decision time with which faces are processed but also the context in which they are encountered. For instance, Davies and Milne (1982) have shown that for unfamiliar faces a significant decrease in recognition accuracy was revealed for changes in context (e.g., background colour), whereas the recognition of familiar faces (famous faces) remained unaffected by such contextual cues.

Although, the situational factors described here are not specifically manipulated in the current research, it is nevertheless important to be aware of their confounding impact on face recognition, because they may considerably influence the results.

Forensic implications

To summarise, various situational factors, such as exposure, delay and orientation can influence the recognition accuracy of familiar and unfamiliar faces, with even stronger impacts on the latter. These findings have important practical implications for forensic settings. For instance, situational factors need to be taken into account by police officers, lawyers, jurors, and judges when determining the quality of witness testimonies or identification evidence, especially since it is nearly always impossible to control for or prevent the influences of situational factors in a naturalistic setting. However, having an understanding of the possible effects allows us to interpret the likely impact on real life scenarios or indeed interpret which might have had an influence in lab based experiments.

4.2.3 Face specific factors

In addition to situational factors, aspects inherent to faces themselves, such as distinctiveness and ethnicity can influence how we process and subsequently remember faces (Hancock et al., 2000). Some of these factors, which appear to be relevant for the following experiments, are discussed below.

Distinctiveness and caricaturing

Various studies have demonstrated that distinctive faces are better remembered than ordinary ones (Valentine & Bruce, 1986a, 1986b; see Hancock et al. 2000 for a review). This holds true for familiar and unfamiliar faces (Bruce & Young, 1998). Valentine and Bruce (1986a) presented participants with faces of work colleagues and asked to rate their familiarity. The faces were previously rated on their distinctiveness by a different sample of participants. Significant negative correlations were obtained between distinctiveness, familiarity ratings and familiarity decision latencies such that both distinctiveness and familiarity contributed to faster recognition of a face as being familiar. In several follow-up experiments Valentine and Bruce (1986b) replicated the recognition advantage of distinctiveness and familiarity with famous faces. However, they failed to find an advantageous effect of distinctiveness on recognition for unfamiliar faces.

Moreover, it was revealed that distinctive jumbled faces took longer to be classified as being 'faces' than nondistinctive jumbled faces, this was true for familiar as well as unfamiliar faces and for famous ones as well as work colleagues. Valentine and Bruce (1986b) demonstrated that participants had more trouble classifying distinctive faces as being 'faces' because they deviate more from a general

prototypical face, which on the other hand facilitates performance on familiarity rating tasks.

Additional support for a facilitating face recognition effect of distinctiveness derives from experimental investigations regarding caricatures. Researchers found that exaggerating distinctive features of faces can have an enhancing effect on recognition. Benson, Davies and Perrett (1994) conducted several experiments investigating the impact of caricature on the perceived likeness of line-drawings of famous people. It was revealed that a high level of caricature was used to increase the perceived likeness of the line-drawings (mean exaggeration was 42% over veridicality, i.e. the correct perception of the image). Furthermore, different faces were exaggerated to different amounts by participants. Particularly, more ordinary faces were exaggerated to a significantly higher degree than more distinctive ones to achieve optimal resemblance. In a subsequent experiment Benson and colleagues (1994) examined whether the optimal caricatures produced during the initial experiment would also have a beneficial effect above the veridicals during a naming task. Indeed, it was found that caricatures were considerably faster and more accurately recognised than veridicals.

A potential explanation for the facilitating effect of distinctiveness and caricature on face recognition was provided by Valentine (1991) with the 'face-space' framework. The face-space theory states that any face can be described by its value along each of a number of dimensions of facial variations (see Figure 4.2 for a simple representation of a two-dimensional face space according to Valentine, 1991). There are two specific models within the face-space framework, the norm-based versus the exemplar-based model (Valentine, 1999). Whereas the norm-based model assumes that each face is encoded in terms of its deviations from a prototypical face in the

centre of the face space, the exemplar-based model postulates that the centre of the face-space is unimportant, but the distance between two faces in space determines their similarity. Contrary, in the norm-based model, the similarity of two faces is described in terms of vectors, which are based on the distance between the two faces and their deviation from the centre (Valentine, 1999). However, both models make similar predictions. Multiple dimensions are required to fully characterise the appearance of a face. Dimensions can be understood as specific facial features, such as nose width or eye shape, or more general attributes, such as age or sex. Ordinary faces will have values on dimensions which are true for many faces, whereas more distinctive faces will have more extreme values which are not shared by many other faces (see Bruce & Young, 1998). Importantly, the model assumes that recognising a face involves making a comparison between the dimensions of this face with dimensions of faces previously encountered. For ordinary faces it is more difficult and may take longer to reach a decision, since there are so many faces in the face space with similar dimensions. On the contrary, distinctive faces will be recognised more easily and faster because only a limited amount of shared values need to be considered (for a comprehensive description of the model see Bruce & Young, 1998). The face-space model can also account for the finding that distinctive jumbled faces take longer to be classified as being 'faces' than nondistinctive jumbled faces (Valentine & Bruce, 1986b).

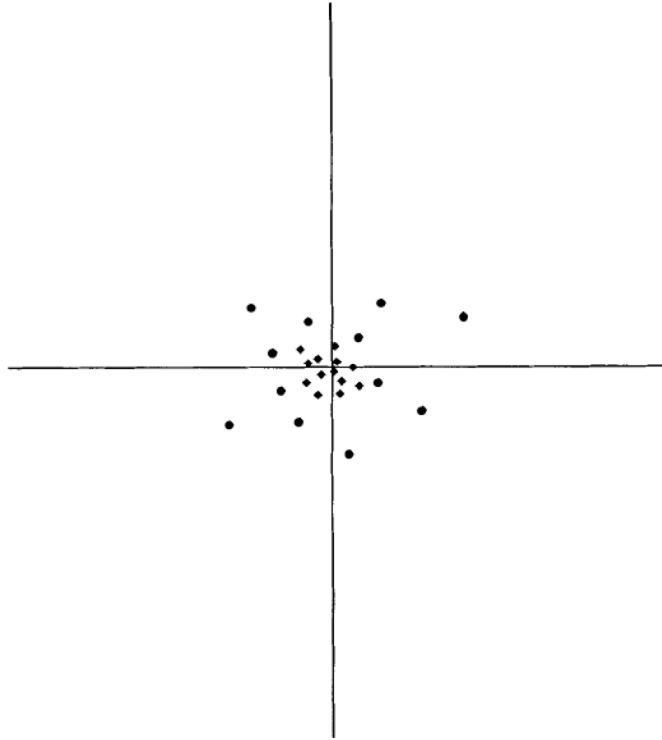


Figure 4.2. *Two-dimensional representation of the exemplar based face-space model by Valentine (1991) (reproduced from Valentine & Endo, 1992). Each point represents a previously encountered face. More ordinary looking faces are plotted towards the centre, whereas more distinctive ones are more distributed. The unlabelled axes could stand for any facial feature attributes, e.g. hair colour, nose breadth, eye shape, etc.*

The studies cited above provide evidence that distinctiveness can have a major impact on face recognition performance. Therefore efforts were made during the present experimental studies to ensure that the employed stimulus material was tested previously on distinctiveness. In addition, during every experiment more than one target face was included, to rule out the possibility that any findings are due to the distinctiveness of one particular face.

External versus internal features

Faces are of course made up of a number of different features. However, not all facial features are remembered equally well. Various studies have found that people rely more on internal facial features (eyes, nose and mouth) than on external ones (hair,

face shape and ears) when recognising familiar faces. However, for unfamiliar faces both types of features appear equally important (Ellis, Shepherd & Davies, 1979; Young, Hay, McWeeny, Flude & Ellis, 1985).

One of the first research teams who investigated the role of each feature category on familiar and unfamiliar face recognition was Ellis et al. (1979). In a series of experiments they presented participants with pictures of famous faces, either in full, as mere inner facial region, or as mere outer region (e.g., silhouette including hair style etc). Participants were required to name them. On average, participants accurately identified 80% of the full faces, 50% of the inner facial regions and 30% of the outer regions. During a follow-up experiment with unknown facial stimuli the authors failed to replicate this finding. Hence, Ellis et al. (1979) concluded that there seems to be a genuine difference in the importance of external and internal features in the recognition of familiar and unfamiliar faces.

Similar findings were obtained by Young et al. (1985) during a matching task using familiar as well as unfamiliar faces. Participants were required to indicate whether two presented faces were either the same or different. For one of the presented faces either the external or the internal features were masked. It was found that familiar faces were matched significantly faster than unfamiliar faces when the internal features were presented. However, no significant difference in reaction time (RT) was obtained between familiar and unfamiliar faces for external features.

Frowd, Bruce, McIntyre and Hancock (2007) tested the importance of the two feature categories in a more applied setting, namely during the construction of facial composites. Composites of unfamiliar faces were created and then evaluated by an independent sample of participants familiar with the people displayed. Participants were presented with either full composites, the internal or the external parts and were

required to sort them according to their resemblance to the target face. Overall, full composites and external features were sorted to a similar accurate degree (33%), whereas internal features were sorted significantly poorer (29.5%). Similar results were revealed with a matching task. Participants had more correct matches on the basis of external composite features than on internal ones. Frowd et al. (2007) argued that the advantageous affect of external features above internal ones on composite recognition might be due to the unfamiliarity of the witness during the construction of the composite (see also Ellis et al, 1979; Young et al., 1985). To test their hypothesis the experiment was replicated, but this time participants created composites of both highly familiar and unfamiliar faces. Contrary to the expectations, familiarity did not have a significant effect. The same results were obtained as during experiment 1 and 2, the quality of the full composites and external parts was better than that of the internal features. The authors concluded that the previous found advantage of external features for unfamiliar faces can not be generalised to the construction of facial composites. A possible reason for this finding may be that in general the quality of internal features presented in facial composite systems is inferior compared to the external ones.

Holistic versus featural processing

Although it was noted earlier that faces are made up of features, numerous studies have demonstrated that the configuration of facial features (the relationships between different facial features) is as important in human face perception and processing as the individual components themselves. Evidence that we perceive faces in a holistic way comes from research conducted by Tanaka and Farah (1993). In their study, participants were asked to learn and remember intact and scrambled faces. Scrambled faces were used because they have the same features or parts as a normal intact face,

but yet we would not expect special face-specific processing abilities to take place. After the learning phase participants engaged in a forced-choice recognition task, during which they had to identify the previously learned facial features. The features were presented either in isolation or in the context of a whole face. In the isolated part test condition, participants were asked to identify 'Larry's' nose and in the full face condition participants were asked 'who is Larry?'. Tanaka and Farah (1993) found, that individuals were able to identify isolated parts from intact faces correctly, but their performance increased when the same parts were tested in the whole face. With scrambled faces, participants were better at identifying the parts tested in isolation than when tested in the context of the whole face. These findings strongly suggest that normal faces are perceived holistically, as a whole object, whereas scrambled faces are perceived in terms of their individual parts.

Another piece of experimental evidence for holistic face processing comes from the composite technique used in an experiment by Young, Hellawell and Hay (1987). These authors divided famous faces horizontally into upper and lower halves and presented those to participants either as a composite or a noncomposite. Composite images constituted of a top-face half joined to the bottom half of a different face. Noncomposites consisted of a top- face half positioned above a bottom-face half in such a way that they did not align. Participants were required to name the face halves as quickly as possible. Overall, participants were very accurate at identifying upper and lower face halves. However, participants responded significantly faster to noncomposites than to composites. Thus, it became more difficult for participants to recognise to whom the face halves actually belonged when they were combined with inadequate other halves. Young and colleagues (1987)

argued that this is because holistic face processing takes place and we cannot simply ignore one face half while evaluating another one.

Further support for holistic face processing derives from the Margaret Thatcher illusion. Thompson (1980) turned the eyes and the mouth in a picture of Margaret Thatcher upside down and noted that when viewed with the entire picture inverted, the expression of the face does not look very different from the original. However, when the picture is presented upright the viewer perceives a particularly grotesque expression. A popular interpretation for this phenomenon is that when we see a face upside down, we do not process it as a whole but rather featurally. Thus, when we merely consider the individual features, they all look acceptable. However, when the face is viewed in a normal orientation, natural holistic face processing is involved and we perceive the face as a whole and immediately recognise that there is something wrong with the positioning of its features.

These findings clearly demonstrate that information pertaining to the normal configuration of facial features is at least as important as the information derived from the features themselves. Contemporary computerised facial composite systems, such as E-FIT and ProFIT incorporate this holistic face processing approach by ensuring that facial features are always presented to the witness in the context of the whole face rather than in isolation (see Chapter 5 for a review of computerised facial composite systems and research evaluating those). This idea was even taken one step further by different research teams and led to the development of purely holistic facial composite system, so called evolutionary systems (Chapter 5 provides a detailed description of one of these evolutionary systems, i.e. EvoFIT).

Forensic implications

The section above reviewed research findings which suggest that similar to situational factors, face-specific factors can have important practical implications in forensic settings. For example, distinctiveness can have a significant impact in an eyewitness situation. Witnesses may be more likely to accurately recognise a distinct looking perpetrator out of a police line-up compared to an ordinary looking one. On the other hand, witnesses might be also more likely to select a distinctive looking foil from a target-present line-up. This might pose challenges to the police during the creation of fair line-ups and countermeasures need to be considered to reduce the impact of distinctiveness. Possible solutions to this problem are line-ups including foils, who replicate the distinctive features or to conceal them. Research has shown that the former appears to be the more effective solution (Zarkadi, Wade & Steward, 2009). Furthermore, the above mentioned research findings demonstrate that humans perceive and process faces in a holistic way rather than in terms of individual features. This knowledge led to the development of contemporary computerised facial composite systems, such as E-FIT and Pro-FIT and novel evolutionary systems, such as EvoFIT.

4.3 Face recall

In contrast to the highly researched area of face recognition only limited research has been conducted in the area of face recall and factors influencing its accuracy and completeness (Shepherd & Ellis, 1996). The existing research can be divided into two broad categories: experimental and archival studies. Although laboratory studies are often criticised as lacking ecological validity and generalisations to real world situations appear problematic, they make it possible to assess omission and

commission errors (Van Koppen & Lochun, 1997) and to control for influencing factors (Sporer, 2001). Both types of research provide important insights into aspects of face memory and can be regarded as complementing each other.

4.3.1 Verbal overshadowing effect

In a series of experiments Schooler and Engstler-Schooler (1990) demonstrated that verbally describing a face can have detrimental impacts on later face recognition. During the first experiment, participants watched videos of a bank robbery. After a 20 minute delay, half of the participants were required to verbally describe the robber's face, the other half engaged in an unrelated filler task. Thereafter, all participants engaged in an identification task. It was found that previously describing the robber's face had a significant negative impact on later recognition accuracy (the target face was identified accurately by 38% of participants who verbally described it and by 64% of participants who engaged in the filler task). Furthermore, the accuracy of the provided facial descriptions was examined by comparing them with feature checklists completed by six independent judges prior to the experiment. A similar evaluation procedure was utilised during Study 2 in this thesis. The features on which most judges agreed were considered as the accurate descriptions. No significant relationships were obtained between the quality of facial features and recognition performance during the identification task. To examine in more depths the cause for the verbal overshadowing effect (VOE), the experiment was replicated with the addition of a third condition, during which participants were required to imagine the target face instead of describing it. Participants in the describing condition performed significantly poorer during the subsequent identification task than participants in the imagination and control condition (recognition accuracy: 27% in the verbalisation condition, 58% in the imagination condition and 60.6% in the control condition). This

finding indicates that recognition interference is caused by verbalising a visual memory of a face rather than merely imaging it.

One possible explanation for the VOE is source monitoring confusion (Dodson, Johnson & Schooler 1997). It might be that people confuse the verbal memory of the face with the visual one and therefore pick the wrong face out of a subsequent recognition test. Dodson and colleagues (1997) tested the source confusion theory by warning some participants prior to the identification task about the detrimental effect of verbally describing the face. Furthermore they included a condition during which participants had to verbally describe the face of their parent instead of the target; this should lead to less source confusion, due to a discrepancy between the likeness of the described face and the target face at identification. Contrary to expectations, no supportive evidence was obtained for the source monitoring confusion theory. Participants who describe the face but received a warning were still significantly less accurate in recognising the target face than control participants. Moreover, participants were less accurate when they described their parents face than when they were in the control condition. On the basis of these findings the authors dismissed the source monitoring confusion theory and argued that a shift from holistic to more featureal processing is more likely to be responsible for the VOE.

In a follow-up experiment, they tested this assumption by presenting participants with either male or female faces. After the completion of a filler task, participants were randomly assigned to three conditions, during which they either had to describe a male face, a female face or no face. Thereafter the recognition test for male and female faces followed. It was found that verbalising either the male or the female faces considerably impaired recognition performance for both male and female

faces compared to the control condition. On the basis of these findings and the previous ones obtained by Schooler and Egnstler-Schooler (1990), Dodson et al. (1997) concluded that it appears more likely that the VOE is the result of a processing shift from a holist approach to a more featural one.

An alternative explanation for the VOE which has reached more and more acceptance was put forward by Clare and Lewandowsky (2004). They argue that VOE is the result of a recognition criterion shift. Verbalising a face can be regarded as a difficult task and it is unlikely that the describer has a suitable reference against which to compare the description. These two factors may contribute to an increased reluctance to pick somebody during an identification task. The criterion shift would lead to a decrease in willingness to pick somebody from a target present (TP) line-up, thus responding with 'not present' and consequently a decrease in hits. However, it would also lead to an increase in correct rejections during target absent (TA) line-ups. During a series of experiments Clare and Lewandosky (2004) tested this theory by including TA and TP line-ups in the standard VOE paradigm. Verbal overshadowing was present with the TP line-ups. Thus, participants made significantly fewer hits when they previously verbalised the description compared to the control group. As predicted, with the TA line-ups, participants in the verbalisation condition were significantly more reluctant to pick somebody from the line-ups, expressed by their significantly lower number of false alarms. These findings have important practical implications. In a real world police setting it is uncertain if the real perpetrator is actually included in the line-up, a shift in recognition criterion might therefore protect innocent suspects from being misidentified (Clare & Lewandosky, 2004). Thus, verbal overshadowing can have both negative as well as positive impacts on face recognition performance depending on the situational circumstances present (see

Meissner & Brigham, 2001 and Meissner, Sporer & Schooler, 2007 for a meta-analysis of the VOE in face identification).

4.3.2 Feature saliency

As mentioned earlier (see section on ‘external vs. internal features’) some facial features might be better remembered than others. In addition, these features might be also more likely recalled on a later occasion (Shepherd & Ellis, 1996). Ellis (1986) termed these features cardinal facial features. Evidence for the existence of cardinal facial features was derived from work by Shepherd, Ellis and Davies (1977), Ellis (1980) and Laughery, Duval and Wogalter (1986). They analysed, independently from another, verbal facial descriptions and found that the most frequently described feature was the hair. Moreover a top to bottom ordering of features was observed, with the hair, eyes and nose being significantly more often mentioned and the mouth and chin region being less likely recalled (see Shepherd & Ellis, 1996 for a more detailed description of these studies).

In addition to laboratory studies, the use of archival data from official police records may offer an additional source of information regarding the quality and quantity of verbal facial descriptions. Sporer (1992) conducted an archival study of US criminal cases and found a feature salience pattern which deviated slightly from the early laboratory research findings (Ellis, 1980; Laughery, Duval & Wogalter, 1986; Shepherd, Ellis & Davies, 1977). In Sporer’s analysis (1992) face shape and skin descriptors were more than twice as likely to be recalled as any other facial features.

A similar research approach was used by Van Koppen and Lochun (1997). They assessed the validity and completeness of offenders’ descriptions by using archival data of convicted robbers in the Netherlands. Overall, the study included 431

cases and 2,299 witness statements. The statements were coded with regard to their accuracy and completeness. This was done by comparing the witness descriptions with police files from the national data base on offenders (Herkenningdienst Systeem). The data base includes written offender descriptions, which are maintained and up-dated on a regular basis each time an offender is arrested. Van Koppen and Lochun (1997) revealed that the most frequently mentioned offender description was sex and the second most prevalent characteristic was height. With the exception of skin colour and hair, facial features were significantly less frequently described. The obtained person descriptions varied considerably in accuracy. The most accurate descriptions were revealed for sex, hair colour and eye shape. Least accurate descriptions were obtained for facial hair. In general, the witnesses' descriptions were more often correct than incorrect. Of the overall 7,754 verifiable offender descriptions 59% were correct, 17% partially correct and 24% were wrong. On the basis of these findings, Van Koppen and Lochun (1997) concluded that person descriptions by witnesses are often vague and include mostly descriptions of general offender characteristics compared to more specific ones, such as descriptions of facial features. The authors argued that featural face descriptions might often be rare because in the real world witnesses frequently have poor views on the perpetrator's face due to variables such as distance, movement and/or disguise.

4.3.3 Forensic implications

The research on face recall is sparse compared to the massive amount of research in the area of face recognition. Two types of research have contributed to the general foundation of knowledge; these are laboratory studies and archival studies. As mentioned earlier both approaches have advantages and disadvantages. Laboratory studies often lack ecological validity and practical forensic implications of findings

remain therefore debatable. On the other hand, they allow to control for influencing factors and to establish causal relations. In contrast, findings from archival studies provide vital insights into the usefulness of forensic evidence, in this case into the ability of witnesses to provide facial descriptions, since they deal with data from real-world cases. Having said this, it is often problematic to assess the accuracy of these data, due to the possibility of false convictions and potential changes to the appearance of the perpetrator over time (Van Koppen & Lochun, 1997). Taken together, the reviewed research indicates that witnesses find it difficult to recall specific facial features and some features are more easily recalled than others. Ironically, the feature which is reported to be most frequently recalled, the hair, is also the one which can be most easily changed by the perpetrator. Findings from the VOE research bare important practical implications to the police, such as that extensive interviewing of the witness or the construction of a facial composite should be avoided prior to the presentation of a mug shot book or the participation in an identification parade. Definitely further research is required to explore in more depth the enhancing and detrimental effects of situational and face-specific factors on face recall.

4.4 General conclusions

Face recognition and recall can play crucial roles in forensic settings, such as during eyewitness identifications or the construction of facial composites. Despite individual differences, the ability to accurately recognise a face or to recall its individual facial features can be influenced by numerous factors, such as situational and face-specific ones. Situational factors include aspects that are associated with the time frame and the surroundings of an event. Face-specific factors on the other hand, are independent

from the actual event, but are inherent to the face itself. How familiar or distinctive an individual face is to the observer can be regarded as face-specific aspect. It is of utmost importance for researchers in the area of memory for faces, to be aware of these influencing factors and their distinct impact on familiar and unfamiliar face recognition, to avoid methodological flaws and the occurrence of confounding factors. In the real-world it is nearly impossible to control these factors and therefore knowledge regarding their impacts can be crucial during the evaluation of eyewitness evidence in criminal investigations and in court. Research investigating these influencing factors has mainly focused on face recognition, whereas face recall has received relatively little attention. This might be due to the fact that people engage much more frequently in tasks involving face recognition than in verbally describing faces (Ellis & Shepherd, 1992). The reviewed research has contributed to the development of influential theoretical frameworks on human face recognition, such as the functional model by Bruce and Young (1986) and the face-space theory by Valentine (1991). On the basis of these frameworks it is possible to make specific predictions about human face processing and to test these empirically. Furthermore, the models provide explanations for basic and more complex research findings, such as the observed dissociation between familiar and unfamiliar face processing and the caricature effect. Overall, the reviewed literature suggests that face recognition and recall research has important theoretical as well as practical implications, with the latter specifically applying to the legal setting. The research comprised in this thesis will contribute to the existing base of knowledge regarding human face processing, specifically by adding to the limited amount of research relating to face processing in individuals with mLD.

Chapter 5

History and development of facial composite systems

This chapter provides an overview of the historical development of facial composite systems, and research evaluating these systems is discussed. First, several early mechanical systems are described and their shortcomings are highlighted. Second, various computerised systems are mentioned and their advantages above older mechanical ones are emphasised. Finally, novel evolutionary systems are described and their potential utility is discussed.

5.1 What is a facial composite?

For the past 100 years, composite images have been used to identify and help capture criminals (Taylor, 2001). As stated in the ACPO Facial Identification Guidelines (2009):

A composite image, as used by the police, is a pictorial likeness produced from the witness's recall of the suspect for the purpose of achieving a 'likeness' of the suspect. The composite image is intended to be an aid to the investigation of crime alongside other corroborative evidence. (p. 9)

Facial composites can be circulated to other police services or the public to gather information about a suspect or to narrow down the range of possible suspects. To date, various facial composite systems have been developed and used to produce composite images. Considerable experimental research has been conducted to evaluate the utility of the systems and to improve both their design and their procedures. In general, there are four categories of composite systems: sketch artists,

mechanical composite systems, computer-based featural, and evolutionary composite systems. The systems which are most frequently used by British and other international law enforcement agencies are described below, together with relevant psychological research evaluating their efficiency.

5.2 Early facial composite systems

5.2.1 Sketch artist

One of the earliest procedures involves a sketch artist who uses his/her artistic skills to create a visual likeness of a perpetrator's face. Although not as popular now, sketch artists are still utilised in this way today by some law enforcement agencies (Davies & Valentine, 2006). Usually, a sketch artist is a person experienced in portraiture who draws a composite image by hand. In general, sketch artists take one to three hours to create a composite drawing (Taylor, 2001). Sketch artists were used in several well-known US crime cases, such as during the Wall Street Bomber case in the early 1920's, the Sam Sheppard murder case in the early 1950's and the Green River Killer case in the 1980's. More recently, a facial composite drawing assisted the police in identifying the perpetrator of the Oklahoma Bombing in 1995 (Taylor, 2001).

Although, the sketching method is one of the oldest procedures used to create facial composite images, there are no international standards available on how to sketch a reliable composite drawing. Consequently, there may be considerable differences in the way individual sketch artists' work. Also, there is a great variability in individual's artistic skills. This causes difficulties when attempting to evaluate the performance of sketch artists empirically (Davies & Valentine, 2006).

5.2.2 Mechanical composite systems

The second category of facial composite system includes mechanical systems such as Identi-Kit and the Photo-FIT system. An advantage of mechanical systems is that they appear to be more standardised and do not require the involvement of a professional sketch artist; instead they can be operated by any trained police operator (Davies & Valentine, 2006).

Identi-Kit

Identi-Kit was established by Hugh MacDonald in 1959 and has been mainly used in the US. The kit consists of acetates with facial features printed on them. Early versions contained acetates with line-drawing on them, whereas later kits include acetates depicting photographic features. These acetates are placed over each other to create a composite image. The original Identi-Kit contained 568 line-drawings of individual facial features (Davies & Valentine, 2006). A claimed benefit of Identi-Kit is that it included a numbering or codification of the feature components to ease the process of reproduction by other operators or police services (Taylor, 2001).

Photo-FIT

Photo-FIT was invented by Jaques Penry in the early 1970s (Penry, 1974) in the UK and was very similar to Identi-Kit. However, instead of acetates, it used templates with photographic facial features printed on them. In contrast to Identi-Kit, where the selected facial features were stacked over each other during the composite construction, Photo-FIT features are assembled next to each other in a facial triangle and are surrounded by a selected face shape (Taylor, 2001). It has been claimed that the process of creating a Photo-FIT is comparable to solving a jigsaw puzzle. Photo-FIT comprised 560 different facial features (Davies & Valentine, 2006).

In addition to the facial feature exemplars, both mechanical composite systems included a collection of accessories, such as spectacles and hats to improve the likeness of the final composite image (Davies & Valentine, 2006). A weakness of both systems is that no specific operating instructions were supplied. However, generally operators were advised to first obtain a verbal facial description from the witness and to subsequently select the most appropriate facial features from the kit. After the assembly of the different facial features, the witness is allowed to further enhance the likeness by replacing or modifying individual features or by changing their position until a satisfactory resemblance emerges (Davies & Valentine, 2006). It is also possible to further enhance the likeness of the composite image by using wax pencils to draw on top of the acetates or templates to add lines and distinctive marks (Taylor, 2001). In addition, Photo-FIT operators were advised to use the book *Looking at faces and remembering them: A guide to facial identification*, written by Penry and Isobel in 1971, to gain further insight and assistance during the composite construction process (Taylor, 2001).

Research evaluating early composite systems

A substantial amount of psychological research has evaluated the effectiveness of mechanical composite systems (e.g. Davies & Christie, 1982; Davies, Ellis & Shepherd, 1978; Christie & Ellis, 1981; Ellis, Davies & Shepherd, 1978; Ellis, Shepherd & Davies, 1975; Gibling & Bennett, 1994; Laughery & Fowler, 1980). Most research focused on the Photo-FIT kit and only limited research investigated the effectiveness of Identi-KIT.

Ellis et al. (1975; 1978) were some of the first researchers who investigated the efficiency of the Photo-FIT kit. In a series of experiments, they assessed the ability of people to reconstruct Photo-FIT composites of unfamiliar faces and to later

identify target faces on the basis of them. The reconstructions were created either with the target face in view or from memory alone. The accuracy of the resultant reconstructions was assessed in two ways. First, the absolute number of correct chosen features was counted. Secondly, a likeness rating was performed by an independent sample of participants. The results indicated that participants had selected more correct features during the target present condition than in the target absent one. The likeness rating showed a similar result. Overall, regardless of whether the target was present or not, participants seemed to have difficulties reconstructing the target composites without mistakes.

A subsequent experiment examined how reliably photographs of faces could be identified on the basis of Photo-FIT composites. Ellis et al. (1975) were also interested in investigating the role of individual differences in participants' ability to create accurate composite images. Therefore, participants who had created good and poor reconstructions in experiment 1 were asked to construct composites in experiment 2. An independent sample of participants evaluated the resultant composites via an identification task. A fairly low hit rate of 12.5% was obtained. On the basis of this finding, Ellis et al. (1975) concluded that people appear to have difficulties in creating composite images using Photo-FIT, even when the target face is present. Interestingly, it was discovered that participants who created good quality reconstructions during experiment 1, also created the more frequently identified composites during experiment 2. Thus, people seem to differ in their individual ability to create accurate facial composites.

Having discovered that individual differences may affect the quality of composite images, Ellis et al. (1978) aimed to identify further influencing factors by manipulating intentional viewing, duration of target exposure, and operator skills.

However, contrary to their expectations, none of these factors had an impact on composite quality.

In a final experiment Ellis et al. (1978) compared the usefulness of a sketching method with those of Photo-FIT, during both a target present and a target absent condition. The quality of the resulting composite images was assessed via a likeness rating. The findings indicated that there was a slight advantage for Photo-FIT during the memory condition. However, during the target present condition, the sketching method was significantly superior. When interpreting these results, it should be noticed, that the sketches were drawn by the participants themselves and not by an experienced sketch artist. Thus, participants' own drawings were better than the composites created with Photo-FIT. This fact raises severe doubts regarding the usefulness of Photo-FIT as an accurate facial composite construction method.

Similar disappointing findings were revealed by Christie and Ellis (1981), who found that participants rated verbal facial descriptions as significantly more accurate than created composites with Photo-FIT.

In order to overcome some of the weaknesses of mechanical composite systems, Gibling and Bennett (1994) tried to improve the quality of composites by using artistic enhancement techniques, such as blanking out unwanted areas and/or adding details. It was revealed that enhanced composites led significantly more often to a correct identification than basic composites. According to Gibling and Bennett (1994), the employment of enhancement techniques can help to overcome the limitations of Photo-FIT observed in previously conducted laboratory studies.

Although a lot of research has examined the effectiveness of Photo-FIT, only limited research evaluated the performance of Identi-Kit. Laughery and Fowler (1980) compared sketch artist performance with that of the Identi-Kit technique. The

resultant composites were evaluated using a likeness rating task. Similar to the findings obtained by Ellis et al. (1975; 1978), it was revealed that sketch artists produced significantly better likenesses than Identi-Kit. Laughery and Fowler (1980) concluded that Identi-Kit might be of limited value to the police.

Summary

The reviewed research emphasises that mechanical systems suffer from various limitations, which might be responsible for their poor performance. First of all, the range of facial features is limited and they lack representativeness of actual facial features. This might be one reason why the sketch artist often outperformed the mechanical systems in several experiments, as the artists can generate an infinite set of facial features (Laughery & Fowler, 1980). A second reason for the superiority of sketch artists above mechanical systems might be that some details, such as shadings and wrinkles, are typically added to sketches but are not available in mechanical composite systems (Laughery & Fowler, 1980). Moreover, the composites created with mechanical systems often have demarcation lines, which might have a negative influence on later identification. Evidence for this stems from a study conducted by Ellis et al. (1978), who found that Photo-FIT composites without these demarcation lines were identified better. The majority of studies did not use any artistic enhancement techniques to improve the appearance of the resulting composites. However, research found that enhancing composites artistically can lead to an increase in correct identifications (Gibling & Bennett, 1994). Moreover, no standardised operator guidelines were available on how to create accurate facial composite images. Therefore, it was difficult to train operators effectively in the use of mechanical composite systems (Davies & Valentine, 2006). Furthermore, during the composite construction process with mechanical systems, witnesses are

continuously exposed to and required to compare facial features. This might have a detrimental and interfering effect on the witnesses' memory and consequently on the quality of the composite (Davies & Christie, 1982). Finally, mechanical systems require participants or witnesses to select facial features in isolation, which are subsequently composed into a whole face. This feature-based approach is in strong contrast with the scientific evidence on how humans naturally perceive and store representations of faces in memory, namely in a more holistic manner (Farah, Wilson, Drain & Tanaka, 1998; Tanaka & Farah, 1993; Tanaka, Farah, Peterson & Rhodes, 2003).

Overall, it can be concluded that mechanical systems do not seem to have advantages over previously used sketching methods. Most studies discovered that the practical utility of mechanical systems under ecologically valid conditions is rather low. It is not surprising that such systems have effectively been superseded by computerised facial composite systems.

5.3 Current facial composite systems

In part as an attempt to overcome the weaknesses of mechanical composite systems, computer-based systems were generated. These systems contain a broader range of facial features and allow easier feature manipulation through graphic packages, like Paint or Photoshop (Davies, Van Der Willik & Morrison, 2000). Moreover, the developers of these systems claim that they encourage a more holistic face processing approach, in that features are viewed in the context of a whole face and never in isolation. This is more concordant with the theoretical assumption on how humans in general perceive and process faces, namely more in a holistic configural way rather than as an accumulation of individual features (Kovera, Penrod, Pappas & Thiel,

1997; Patterson & Baddeley, 1977; see Chapter 4 for a more in depth review regarding holistic face processing). Computer-based composite systems include featural systems such as Mac-a-Mug Pro, Faces, E-FIT and ProFIT, as well as evolutionary systems such as EvoFIT.

5.3.1 Computer-based featural composite systems

Mac-a-Mug Pro

In 1986 Shaheerazam designed Mac-a-Mug Pro for the use on Macintosh computers. It consists of an extensive database of line-drawn facial features and in addition comprises special editing tools, such as erasers, pencils and brushes. With the assistance of these editing utensils the operator can adjust the facial likeness by adding age lines, cutting hair, adjusting the skin colour, decreasing or increasing the size of features and changing the position of them. All of the facial features are independent of each other and can be moved and modified without having an effect on other parts of the face. Moreover, it is possible to transfer the resulting composites into other graphic programs, such as MacPaint, to further manipulate the appearance. Mac-a-Mug Pro also contains an extensive range of accessorize, such as hats, moustaches and beards, spectacles and sideburns. Hard copies of the resulting composites can be printed and additional information regarding the construction date and time and a written description of the suspect can be added. It offers the opportunity to store composites and features in a digital form, thereby allowing easier retrieval, storage and production of hard copies without considerably altering the original composite image (Kovera et al., 1997).

E-FIT

Another computer-based system widely used by British and European law enforcement agencies is E-FIT (Electronic Facial Identification Technique). Now owned by VisionMetric Limited (2008), it was originally developed by the Home Office and researchers from the Psychology Department at Aberdeen University. A potential advantage of E-FIT over Mac-a-Mug Pro is that it uses features of photographic quality (Davies & Valentine, 2006).

In general, the following method is recommended when creating a composite image with E-FIT. To start with, the operator should explain the purpose of the appointment and the witness will be familiarized with the E-FIT system. Thereafter, the operator will conduct a CI with the witness to obtain a detailed and accurate facial description. The facial descriptors provided by the witness during the CI, will then be translated into the Aberdeen Index³ and entered into facial description boxes by the operator. The entered information drives an algorithm based on fuzzy logic that automatically selects the best fitting features from the database. The resultant E-FIT is then presented to the witness who is allowed to make changes to features by changing their size or position or by scrolling through alternative features within the context of the whole face. When a sufficient likeness emerges further fine-grained changes can be made by the use of standard drawing packages, like Paint or Photoshop. The E-FIT construction process is completed when the witness expresses that he/she is satisfied with the likeness of the composite image (Davies & Valentine, 2006).

³ The Aberdeen Index is a system created by the Psychology Department, University of Aberdeen, Scotland, to categorise facial features.

ProFIT

The ProFIT system, formerly known as CD-Fit, is very similar to E-FIT and uses the same composite construction procedure (Frowd et al., 2005). It was marketed by a UK company called ZEDA in 1999.

FACES

FACES is the US equivalent to E-FIT. It contains photographic facial features and is marketed by a company called IQ Biometrix. Although similar in its design, it is less flexible in its operations, such as resizing and repositioning features. Another shortcoming of this composite system is that the brightness and contrast of features cannot be modified. Benefits of FACES, according to the sellers, are that no special training is required to operate the system and it is much more affordable than other computerised systems, e.g. E-FIT (Frowd et al., 2005). A survey study conducted by McQuiston-Surrett, Topp and Malpass (2006) found that FACES is very popular. It was the second most frequently used computerised composite system in the US (Identi-KIT was the most frequently used system) at the time.

Research evaluating computer-based composite systems

Several empirical studies have evaluated the usefulness of computer-based composite systems and have compared their performance with that of mechanical systems (Brace et al., 2006; Cutler, Stocklein & Penrod, 1988; Davies et al., 2000; Koehn & Fisher, 1997; Kovera et al., 1997). The two systems which have been subject to most extensive psychological research are Mac-a-Mug Pro and E-FIT (Davies & Valentine, 2006). The experimental studies and their results are described and discussed below.

Cutler et al. (1988) carried out one of the earliest empirical studies testing the utility of Mac-a-Mug Pro to create realistic and accurate facial composites. They

asked participants to recognise photographs of target persons. Participants were allocated to one of three conditions. One group based their recognitions on Mac-a-Mug Pro composites that were in view during the recognition test. The second group based the recognitions on their memory for Mac-a-Mug Pro composites. And the third group served as a control group and based their recognitions on memory for photographs of the actual target persons. All composites were created by experienced operators from photographs. Overall recognition performance was good, 68% of the target photographs were correctly recognised. Recognition performance was significantly better during composite present judgments than during the from-memory judgments. There was no significant difference between recognition performances during the memory for composites condition and the memory for photographs condition. On the basis of these findings, the researcher concluded that Mac-a-Mug Pro can produce recognisable composites by skilled operators. However, it should be noted that the operators created the composites with the target photographs in view, which is not an ecologically valid condition.

Less favourable results were acquired by Koehn and Fisher (1997). They examined the usefulness of Mac-a-Mug Pro with three different evaluation tasks: a feature matching task, a likeness rating and an identification task. The results of all three evaluation tasks demonstrated that the composites created with Mac-a-Mug Pro were of very low quality. During the feature-match method, of the 46 composites, 25 obtained zero matching features, 19 had one matching feature and two received two matching features. The maximum number of possible matching features was 11. A similar disappointing result was obtained during the likeness rating. Almost all ratings were very low: on a 10-point scale, 57% of the composites received a rating of one and 12% of two. During the identification task, merely 7% of the participants

correctly identified the targets out of the line-ups. Thus, overall, the composites were rated as extremely poor representations of the target photographs.

Similar results were obtained by Kovera et al. (1997). Familiar faces served as targets, such as photographs of classmates and teachers of the participants. These stimuli were considered as forensically relevant and ecologically valid, since in the real world it is more likely that people who are familiar with the perpetrator will recognise him/her from the composite image. The accuracy of the composites was assessed by a familiarity rating task and a naming task. The study failed to provide evidence for the usefulness of Mac-a-Mug Pro. Nearly no recognisable composites were created and participants' recognition, as well as, naming rates were below chance performance. Only three composites were correctly named whereas 167 were mistakenly named. However, it should be noted, that operators were first-year University students, who had not received appropriate training in the use of the Mac-a-Mug Pro system. The training they got prior to the composite construction consisted of the Mac-a-Mug manual and a brief explanation on the operation of the system. The absence of sufficient training in the use of Mac-a-Mug Pro might have had a detrimental effect on the quality of the resultant composite images.

In a more recent study carried out by Davies et al. (2000) the performance of E-FIT was compared with that of Photo-FIT. Every participant constructed composites with both systems. Composites were first constructed from memory alone; the target-absent condition. After a sufficient likeness emerged the target face was reintroduced and the composite was further amended; the target-present condition. Targets included faces that were both familiar and unfamiliar to each participant. Thus, each participant produced four composites: two E-FIT composites, one from memory and one from photo, and two Photo-FIT composites, one from memory and

one from photo. During the evaluation phase, a different sample of participants completed a naming task, a matching task and a familiarity rating task. The results of this comparative study provided little support for the benefits of computer-driven systems over mechanical ones. On all evaluation tasks E-FIT was only significantly superior during the target present condition. On the naming task and the matching task, this superiority effect was restricted to familiar faces. Thus, during the more ecologically valid condition, E-FIT performed not better than Photo-FIT. On the basis of these findings, Davies et al. (2000) concluded that computer-based systems do not perform better than older mechanical systems and this casts doubts on their overall usefulness.

One might argue that the findings of this study should not be generalised because of several methodological limitations. First of all, no CI was carried out, although this is the recommended interview technique to create a facial composite image. Second, a time limit of 20 minutes was used to create a composite, which is rather short, given the fact that it might take up to one and a half hour to create a composite (Taylor, 2001). No artistic enhancement techniques were used in the end, although Gibling and Bennett (1994) found that artistic enhancement can significantly increase the accuracy of facial composites. Participants created more than one composite during one session which may have influenced their performance, due to concentration difficulties and cognitive overload. Finally, participants created the composites by themselves without the help of an experienced operator and the training participants received prior to the composite construction can be regarded as insufficient.

Brace et al. (2006) evaluated E-FIT under more ecologically valid conditions. Participants created E-FIT composites with the target photograph in view and from

memory alone. To further investigate witness-operator communication and the impact of this on later composite quality, composites were created either directly by the operator or together with a witness as a describer. Target stimuli were pictures of famous faces. The quality of the resulting composites was assessed with a likeness rating and a naming task. Overall, participants were able to correctly name a substantial amount of composites. 66% of all composites were correctly identified by at least one person. Composites were rated and identified significantly better when they were created by the operator alone than with a describer. The presentation mode (from memory vs. from photo) had only a significant impact on quality when composites were created together with a describer; better composites were created during the photo condition than during the memory condition. These findings indicate that the quality of the resultant composites might be impaired by the translation of the verbal description by the operator and not due to difficulties with the system itself. According to Brace et al. (2006), these findings demonstrate that E-FIT, compared to earlier mechanical composite systems, is a more sensitive system with which accurate composite images can be created.

Summary

Overall, research has revealed mixed findings regarding the superiority of computerised featural composite systems to mechanical ones. Several studies revealed rather discouraging findings with regard to the performance of computer-based systems, such as Mac-a-Mug Pro and E-FIT (Davies, et al., 2000; Koehn & Fisher, 1997; Kovera et al., 1997). However, it should be noted that the applied procedure in most of these studies was far from being flawless and often not ecologically valid. One major shortcoming of these studies was that composites were generated by the participants themselves, rather than together with a trained operator. Currently, the

main provider of facial identification training to police services in the UK is Visionmetric at Kent University. Usually their training courses for police operators take two weeks to train officers sufficiently in the application of the composite software⁴, such as E-FIT. The circumstance that in the majority of studies (Cutler et al., 1988; Davies, et al., 2000; Koehn & Fisher, 1997; Kovera et al., 1997) undergraduate university students, with no prior facial composite experience let alone training, constructed the composite images may have had a detrimental effect on the resultant composite quality.

Research including a more ecologically valid procedure (Brace et al., 2006) obtained more encouraging results regarding the performance of computer-based featural composite systems. They demonstrated that arising difficulties with these systems might be rather attributed to problems with the operator-witness communication than to the systems itself.

One of the major claimed advantages of computer-based featural composite systems, such as Mac-a-Mug Pro, E-FIT and ProFIT, is that they should encourage holistic face processing. However, one could argue that they are not entirely holistic composite systems at all. The witness still needs to describe and then work on individual facial features. A composite system which is regarded as a purely holistic one is EvoFIT.

5.3.2 Evolutionary composite systems

The facial composite systems introduced so far all require the witness to divide the face into, concentrate on, and make decisions about individual facial features. This is a particularly demanding task, especially since research has shown that faces are

⁴ Information obtained from the Vision metric webpage, section Training, http://www.visionmetric.com/index.php?option=com_content&task=view&id=36&Itemid=53 on the 3rd September, 2010.

naturally processed in a more holistic way (Tanaka & Farah, 1993; Thompson, 1980; Young et al., 1987). To date, three holistic software programs have been developed, independently from each other, by different Universities. Researchers at Stirling University, Newcastle University and the University of Central Lancashire have developed EvoFIT (Hancock, 2000; Frowd, Hancock & Carson, 2004), Eigen-fit (Gibson, Pallares Bejarano & Somolon, 2003) was developed by the University of Kent, and ID by the University of Cape Town (Tredoux, Nunez, Oxtoby & Prag, 2006). All three software programs use a genetic algorithm to produce recognisable likenesses of faces. The system which has been subject to most empirical research and which is also used during one of the experimental studies in this thesis is EvoFIT and will therefore be discussed below in more detail.

EvoFIT

EvoFIT uses a completely different approach to generate facial composites than the mechanical or computer-based composite systems mentioned above (Frowd et al., 2004). The system is based on the notion that humans tend to have difficulties with describing faces, but are often much better at recognising previously seen faces (Frowd et al., 2005). The EvoFIT system combines so called “eigenfaces” to achieve a high-quality facial likeness in the end (Frowd et al., 2004). The composite construction process begins with the selection of the most suitable hairstyle. Following this, randomly generated faces are presented to the witness. The witness is asked to select several face shapes and face textures (internal facial features) which resemble the target face most. The chosen shapes and textures are then bred together via a combination of an evolutionary algorithm (EA) and principal component analysis (PCA) and a new generation of faces is created. This selection and breeding process is then repeated through a number of generations until gradually an acceptable

likeness is achieved (Frowd et al., 2004). Finally, the size and location of individual features can be manipulated according to the witness's preference. The Lancashire police is among the first police service in the UK to put the EvoFIT system into practice and has achieved already successes with it. In a very recent case in 2008, EvoFIT helped the Lancashire police to convict a sex attacker in Blackpool. The 18 year old perpetrator received a seven year sentence for attempted rape of a school girl. The victim was able to create a very good likeness of the attacker with EvoFIT and within hours the perpetrator was found (Abm United Kingdom, Ltd, 2008). Another crucial benefit of EvoFIT might be that no operator translation of the witnesses' description is required anymore, which might have a detrimental effect on composites' quality, as discovered by Brace et al. (2006).

Research evaluating EvoFIT

One of the first studies assessing the usefulness of EvoFIT as an alternative and improvement to previous facial composite systems was carried out by Frowd et al. (2004). During a series of experiments, participants were asked to create composites with E-FIT and EvoFIT of famous faces that were unfamiliar to those creating the composite images. The composites were finally evaluated by a naming task. Experiment 1 included no delay between the presentation of the target face and the subsequent composite construction from memory. A significant difference was revealed for composite system, with E-FIT (16%) producing composites named more often than composites produced with EvoFIT (7%). During a subsequent experiment, a two-day delay was included between the presentation of the target face and the later composite construction. The performance of E-FIT, EvoFIT and ProFIT was examined and compared. The remaining procedure mirrored the one employed during experiment 1. This time, EvoFIT composites were named significantly better (3.6%)

than composites created with ProFIT (1.3%) and E-FIT (0%). The results suggest that under more ecologically valid conditions (e.g. a longer target delay), EvoFIT performs at least as good as computer-based systems.

In a subsequent study, Frowd et al. (2005a) compared the performance of EvoFIT with five other facial composite systems: E-FIT, ProFIT, sketch and Photo-FIT. An ecologically valid procedure was applied by using famous faces as target stimuli, which were unfamiliar to the participants who constructed the composites but familiar to those participants who evaluated the composites subsequently. In addition, a four hour delay between exposure of the target face and the actual composite construction was included. The evaluation phase included a spontaneous naming task and a sorting task. During the sorting task, participants were required to sort all composites into piles, given the target photographs as references. The findings confirmed the results previously obtained by Frowd et al. (2004) in experiment 1, who demonstrated that E-FIT performed better than EvoFIT. For the naming task, a significant effect for the construction technique was obtained, with E-FIT outperforming all other techniques, except ProFIT. Overall, the mean naming rates were quite low. The mean naming rate for composites created with E-FIT was 19%, 17% for ProFIT composites, 9% for composites created by a sketch artist and 6% for Photo-FIT composites. EvoFIT composites had the lowest mean naming rate of 2%. The findings from the sorting task were in agreement with the ones obtained during the naming task. A significant difference between composite construction techniques was revealed. Composites from E-FIT, ProFIT and sketches were sorted to an accuracy level of around 70-80%, whereas composites created with EvoFIT and Photo-FIT were sorted to an accuracy of approximately 50%. However, it should be considered that at this point in time EvoFIT was in its early stages of development

and therefore may have showed weaknesses. According to the authors, operators found it for example very difficult to operate the software, which may had a detrimental impact on resultant composite quality. E-FIT and ProFIT, which both belong to the category of computerised composite systems performed at an equivalent level and performed better than mechanical systems and sketches. This is in agreement with the results obtained by Brace et al. (2006) and suggests that computerised systems can create qualitative good facial likeness when tested under ecologically valid conditions and that they comprise at least some advantages over older mechanical systems.

During a further study conducted by Frowd et al. (2005b), the performance of EvoFIT was compared with those of E-FIT, ProFIT, sketch, and FACES. Famous faces were used as target stimuli and a two day delay was included between target exposure and composite construction. The procedure closely followed the one applied by the police. To evaluate the quality of the resultant composite images a naming task, a sorting task and an identification task were employed. Overall, a low composite naming rate of only 3% was revealed, indicating again that facial composite systems seem to face serious problems. The sketch technique produced composites which were significantly more frequently named as those created with E-FIT and ProFIT. No other significant differences were obtained. Thus, during the naming task, sketches outperformed all other systems, with a naming rate of 8%, this was followed by EvoFIT (3.6%) and FACES (3.2%), E-FIT and ProFIT performed worst with a naming rate of less than 2%. An overall accuracy of 42% was observed for the sorting task. Again sketch performed significantly better than all other systems (54%), followed by E-FIT (42.5%), ProFIT (40.6%), EvoFIT (38.8%) and Faces (35.0%). During the identification task, an overall correct identification rate of 42.1% was

obtained. E-FIT performed significantly better than all other techniques (60%), except sketch (47%). ProFIT composites were correctly identified 41% of the time, and Faces and EvoFIT composites around 30%. These findings are in agreement with the ones obtained during previous studies (Frowd et al., 2004; 2005) and suggest that composite quality was very poor in general. However, it should be recognised that performance of EvoFIT had improved, compared to the results revealed during previous studies (e.g. Frowd et al., 2004). According to the authors (2005), this might be due to technical improvements of the EvoFIT software tools.

More favourable results for the EvoFIT system were obtained in a study conducted by Frowd et al. (2007). The performance of two different versions of ProFIT (serial and parallel) was compared with that of EvoFIT. Familiar faces were used as targets and a two-day target delay was employed. The resultant composites were subsequently evaluated via a sorting task. Surprisingly no difference was found in composite quality between the two systems. During a follow-up study, a different evaluation task was employed, i.e. a naming task. The results revealed that naming rates were significantly higher for composites created with EvoFIT (8.5%) than ProFIT composites (3.7%). According to the authors (2007), the evaluation instrument (sorting task vs. naming task) might have been responsible for the discrepancy in findings. They claim that a feature-based evaluation task, such as a sorting task, might be not appropriate when assessing the performance of EvoFIT, which is a holistic facial composite system.

Recently, Frowd et al. (2010) evaluated the latest version of the EvoFIT system, which included two enhancements, the blur and the holistic tool, and compared it with ProFIT. The blur enhancement is based on research findings (Ellis, Shepherd & Davies, 1979; Young, Hay, McWeeny, Fude & Ellis, 1985) suggesting

that for the recognition of familiar faces the internal facial features are more important than the external ones, for unfamiliar faces both types of features are equally important. Therefore, the authors (2010) blurred the external features during part of the composite construction process, so that people focus more on the internal features during the creation of the composite, which should increase subsequent composite recognition. The second enhancement technique was the holistic tool. This tool allowed participants to make further changes to the composite image by changing it on several dimensions, such as age, face weight and attractiveness (a detailed description of the development of this tool can be found in Frowd et al., 2006). The study included a two-day delay and targets were famous faces. The resultant composites were evaluated with a naming task. It was revealed that EvoFIT produced significantly better composites than ProFIT. On average, EvoFIT composites were correctly named 24.5% whereas ProFIT composites were only correctly named 4.2% of the time. On the basis of these findings, Frowd et al. (2010) claimed that the improved EvoFIT system is superior to previous computer-based systems and that it appears to be a valuable technique in suspect identification.

Summary

The scientific evidence regarding the efficiency of EvoFIT seems to be mixed. However, it should be kept in mind that EvoFIT has proven already its practical utility and relevance, since it successfully assisted the police in solving a criminal investigation. During the past 10 years, EvoFIT has undergone several improvements regarding its operating tools and the representation of facial features, such as the eyes and the hair styles (Frowd et al., 2004). Furthermore, additional enhancement techniques were developed and included in the EvoFIT package, such as the blur effect and the holistic tool (Frowd et al., 2010). Recent studies demonstrated that

together with these improvements, EvoFIT appears to perform at least as good as, if not superior to, computerised facial composite systems (Frowd et al., 2007, 2010). Another considerable advantage of the EvoFIT system appears to be that it does not depend to a large extent on witness-operator communication. Although, the operator assists the witness by providing instructions and entering the witness's responses manually during the composite construction process, most of the time the witness works directly with the system itself. This forms a substantial benefit for EvoFIT, since research has shown that composite quality can be limited by the fact that witnesses have to work closely together with operators to construct composites with previous computerised systems, such as E-FIT (Brace et al., 2006). Finally, EvoFIT might be even more suitable for specific groups of witnesses, such as children and people with LD. Those witnesses might have particular difficulties verbalising a description of a perpetrators' face and would benefit from the circumstance that during the composite construction with EvoFIT no verbal description is necessarily required.

5.4 Discussion

The reviewed literature in this chapter demonstrates that facial composite systems have undergone dramatic developmental changes during the past decades. Early attempts were quite ingenious, though limited in their applications and resultant composite images were often of very poor quality. Some of the undertaken changes were based on practical problems with previous mechanical systems, such as the failure to further enhance the composite images by adding more details. Modern computer-driven systems offer the possibility to further enhance composites by the use of graphic packages such as Paint or Photoshop. This appears to be a major

advantage compared to earlier systems, since research has demonstrated that artistic enhancement techniques can significantly increase composite quality (Gibling & Bennett, 1994). Other modifications derived from new theoretical insights in how humans perceive and process faces. Numerous studies have shown that humans naturally encode, store and recognise faces in a more holistic/configural manner (Tanaka & Farah, 1993; Thompson, 1980; Young et al., 1987). This stands in strong contrast to the composite construction process with early mechanical systems. Modern systems, such as E-FIT and EvoFIT encourage more holistic face processing. Contemporary composite systems have tried to overcome the weaknesses of early mechanical systems and various studies have proven their superior performance (Brace et al., 2006; Cutler et al., 1988; Frowd et al., 2005a).

It is not only composite systems themselves that have undergone major changes during the preceding years, but also the experimental procedures with which their utility has been assessed. On one hand researchers have tried to implement more ecologically valid composite construction conditions, by including forensically relevant target delays (Frowd et al., 2004, 2005a, 2005b, 2007, 2010; Koehn & Fisher, 1997) and trained operators (Brace et al., 2006). On the other hand, the evaluation tasks used to assess the resultant composites' quality got more realistic as well, by using famous faces as stimuli which were unknown to the person who constructed the composite, but familiar to those who evaluated it. As a result, spontaneous naming could be used as evaluation task (Brace et al., 2006; Frowd et al., 2004, 2005a, 2005b, 2007, 2010). This method can be regarded as highly ecologically valid, since witnesses who construct composites do not know the identity of the perpetrator but composites are more likely to be identified later by somebody who is familiar with the perpetrator.

Although these changes in methodology did increase the validity as well as the reliability of facial composite studies, they also made it problematic to compare research findings across studies and to generate universal conclusions. However, despite the named difficulties, in conclusion it seems reasonable to claim that facial composite systems have developed in a positive way and current systems, such as E-FIT and EvoFIT appear to be valuable tools which can assist the police in criminal investigations. Nevertheless, further research is needed which investigates the newer computerised and evolutionary composite systems under different testing conditions (differences in target delay, differences in viewing conditions, impact of operator's behaviours, etc.) and with different participant populations, such as vulnerable witnesses. As such, one of the major aims of this PhD thesis is to investigate the performance of contemporary facial composite systems, like E-FIT and EvoFIT, with mLD witnesses.

Chapter 6

Study 1: A survey of facial composite operators

The present chapter describes a survey study, which explores experiences and opinions of UK facial composite operators with witnesses with and without LD. The purpose of the survey study is to identify current police practice with LD witnesses and any practical problems facial composite operators face when dealing with witnesses with LD.

6.1 Introduction

In 2003, the ACPO Working Group for Facial Identification stated in their own guidelines for police operators: “Serious consideration should be given to the potential evidential value and accuracy of the recognition and recall factors from: very young or old witnesses, witnesses who are mentally impaired, and witnesses impaired by alcohol or drugs.” (ACPO, 2003, p.10). Such a statement might have severe consequences for witnesses with LD, since it may call their ability into question to construct accurate facial composite images and might bias operators’ attitudes towards their eyewitness abilities in a negative way. A comprehensive search of relevant literature suggests that this statement is not fully based on scientific research findings. Until now, no empirical study has investigated the abilities of witnesses with LD to construct facial composite images.

Interestingly, the most recent Facial Identification Guidance document, which was released in 2009, no longer includes the above mentioned discriminatory statement. However, it does not contain any other specific guidance for police operators when they engage with LD witnesses. Furthermore, it does not refer to any

training that is available to operators providing advice on interviewing witnesses with LD when constructing a facial composite image. The only guidance offered in this document is that a supporter should be available to offer assistance to the LD witness during the facial composite construction processes. This guidance appears to be quite meagre; since no additional recommendations are provided with respect to what form this support should take. The lack of research regarding LD witnesses' abilities to create facial composite images implies that there is no specific guidance available at all. Therefore, the aims of the current study are:

- 1) To identify current practices utilised by police operators during the composite construction.
- 2) To explore operators' opinions about and experiences with witnesses with LD.
- 3) To identify practical problems operators' might face when engaging with LD witnesses.

No hypotheses were put forward for this study, since the aim of the study was to explore experiences and the opinions of police operators.

6.2 Method

6.2.1 Participants

Copies of the survey were directed to a sample of UK law enforcement agencies. The survey was sent to Chief Constables who were asked to distribute it amongst police operators in his/her agency. Unfortunately, this procedure resulted in a very low response rate. Therefore, copies of the survey were also sent to the current distributor of E-FIT (Dr Christopher Solomon) and EvoFIT (Dr Charlie Frowd), who circulated it further. Overall, 17 operators completed the questionnaires. On average the

operators had 4.5 years experience in constructing facial composites, ranging from one month to 17 years.

6.2.2 Survey instrument and procedure

A 27-item survey was developed which addressed issues related to facial composite systems and witnesses with LD (see Appendix 1 for a copy of the survey). The survey included an introductory page explaining the motivation for the study and introducing the author. It contained five open-ended questions, eight multiple-choice questions and 14 Likert-rating scales. The questions can be divided into three broad categories: personal questions (including questions regarding which facial composite system operators are working with, how much and what type of training they had participated in and an estimate of the number of composite images generated); questions regarding experiences with and opinions about composite construction in general (including questions such as: “How easy do you find it to select facial features from the feature database? In general, how detailed are facial descriptions provided by witnesses?”); and questions regarding experiences with and opinions about witnesses with LD (including questions like: “How detailed are verbal facial descriptions provided by witnesses with LD compared to witnesses without LD? Are there any specific guidelines that can be referred to when generating a facial composite image with the assistance of a witness with LD?”). At the end of the survey, there was an opportunity for the respondent to provide suggestions, questions and critique. The survey was a self-administered questionnaire. Prior to distribution, the survey was evaluated by an E-FIT operator from the Grampian Police to ensure that the questions were well understood, the provided answer options were meaningful and the wording of the questions was in agreement with the terminology used by the sample population. The survey was sent to the addressees with a return postage-paid envelope and a covering

letter. The covering letter stated that the survey should be forwarded to a person responsible for the construction of facial composites within the agency. Furthermore, it contained background details, the purpose of the survey was outlined, the desired return date was provided and the contact details of the person in charge of the study were given.

6.3 Results

The results section reflects the organisation of the questions asked in the survey. Response rates and findings were calculated into percentages. The total number of operators that answered the specific questions are always listed at the beginning of each section ($n = x$). Due to the low response rate, it was not possible to generate inferential statistics and the analysis is therefore limited to a descriptive analysis. This is not ideal but a descriptive analysis still affords an understanding of the general trends present in the data.

6.3.1 Personal details

Which facial composite system do you have experience with?

($n = 17$) The data indicated that UK police operators have experience with a variety of different facial composite systems. E-FIT is the composite system most operators have experience with (41%), followed by EvoFIT and ProFIT (14% each) (see Figure 6.1). However, it should be noted that this finding might be somewhat biased, since one method of distributing the surveys was through the manufacturers of E-FIT and EvoFIT.

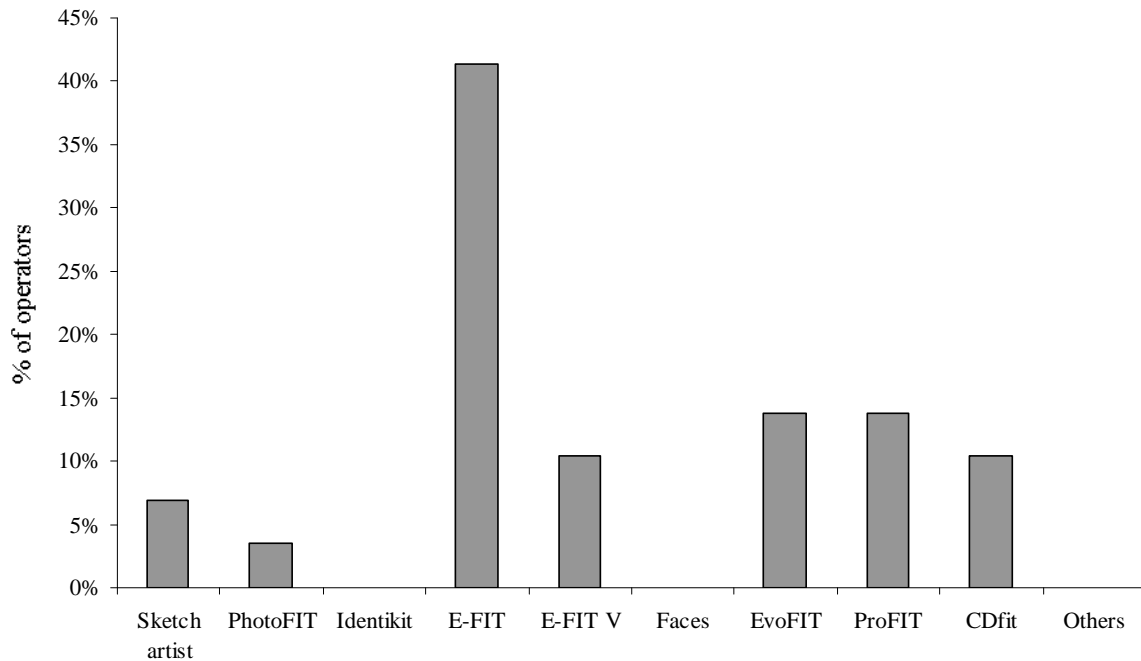


Figure 6.1. *Percentage of facial composite systems operators have experience with.*

Which facial composite system are you currently using?

(n = 17) The majority of operators currently work with E-FIT (63%). The second most currently used composite system was ProFIT (16%), followed by sketch artist and EvoFIT (11% each) (see Figure 6.2). Again the findings should be considered with caution, since they might be somewhat biased due to the distribution method applied.

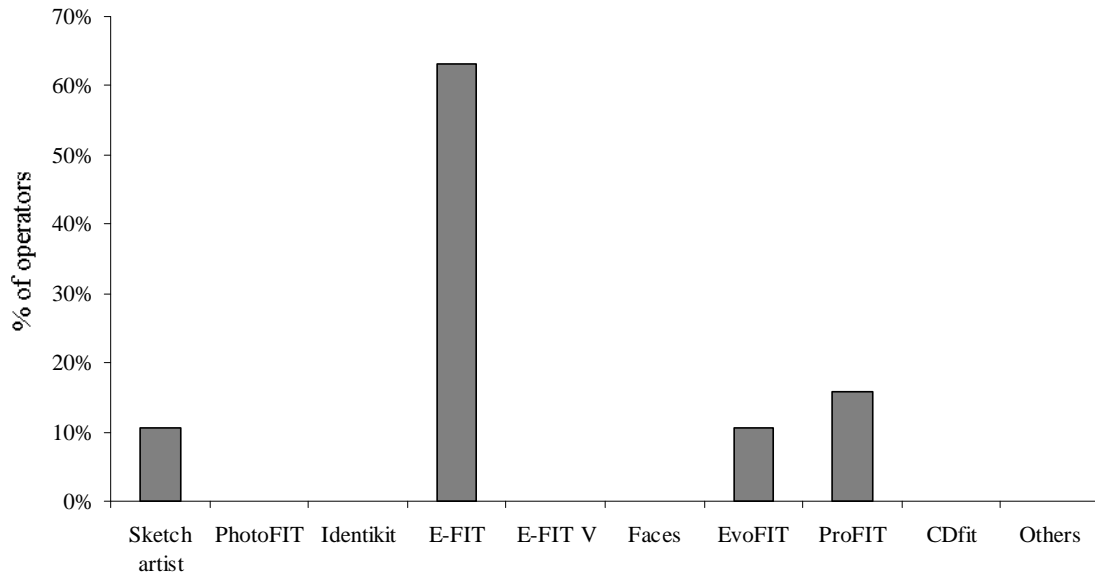


Figure 6.2. *Percentage of facial composite systems operators are currently working with.*

How many years have you been working with the composite system indicated in Question 2?

(n = 15) The data indicated that operators varied considerably in their background and experience. Responses ranged from one month to 17 years. The mean length of time operators had worked with the facial composite system they currently used was just over 5 years.

Did you receive training in the use of the facial composite system indicated in Question 2?

(n = 17) All operators answered that they had received training in the use of the facial composite system they currently worked with.

What kind of training did you receive?

(n = 17) Operators received training from a variety of sources. Most operators received training from the National Police Training Centre in Durham (36%). 18% of

the respondents received training from another officer in their police station or precinct. A further 18% received training by the licence holder of the composite software provider and another 18% were trained by another source (responses included answers such as: course at art school and training course delivered by Dr Charlie Frowd, (University of Central Lancashire). Finally, 9% of the respondents reported that they had obtained training from the Scottish Police College (see Figure 6.3).

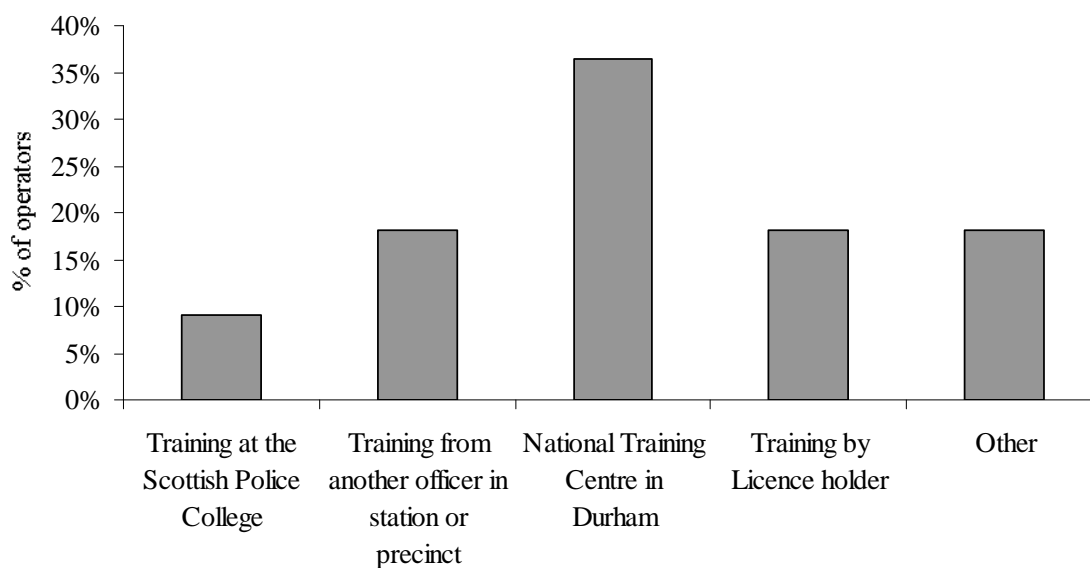


Figure 6.3. *Sources which provided training to operators.*

How many months ago was the training?

(n = 13) The training police operators received was on average 4.5 years ago. Responses ranged from 6 months to 11 years.

Please estimate how many composites you personally have generated during the last two years?

(n = 17) As with previous questions, there was a wide range of responses. Some operators stated that they have constructed only one composite during the last two

years, while others reported to have generated up to 150 composites. The mean number of composites operators constructed during the last two years was 38.88. The majority of operators were constructing 20 composites during the last 2 years.

Please estimate how many composites your department has generated during the last two years?

(n = 17) According to the operators, the mean number of facial composites generated by the different departments during the last two years was 44.12. The majority of departments have generated around 25.5 composites during the past two years.

Summary of main findings

The survey data indicated that UK police operators have experiences with a variety of different facial composite systems. However, E-FIT appears to be the most popular facial composite system. The majority of operators, who have completed the questionnaire, have experience with and also currently work with E-FIT. It should be noted that this is also the composite system, which will be further examined with mLD individuals in this thesis. Furthermore, the data suggested that operators varied considerably in their background and experience. Some operators have had 17 years experience as a police composite operator, whereas others merely have had one month experience. On average, operators had 4.5 years experience in constructing facial composites. All operators received some kind of training in the use of facial composite systems. The specific source of the training varied considerably, but the majority of operators received training from the National Police Training Centre in Durham. The training police operators received was on average 4.5 years ago. Most operators have constructed 20 composites during the last 2 years, but as with previous questions, there was a wide range of responses.

6.3.2 General experiences with and opinions about composite construction

To identify current standard practice and potential difficulties during the composite construction process, operators were asked to answer a series of multiple choice questions regarding their own experiences with and knowledge about the different composite construction phases.

How easy do you find it to select facial features recalled by the witness during the composite construction phase?

(n = 17) Most operators responded that they find it **easy** to select facial features recalled by the witness during the composite construction (59%). 29% of the operators answered that they find this task **very easy**. Only 12% reported that they experience this task as **difficult** and none answered that they experience it as a **very difficult**.

How much does the construction of a facial composite depend on the language abilities of the witness?

(n = 17) Most operators believed that the language abilities of the witness play only a trivial role during the composite construction. 47% stated that the composite image depends only **little** on the language abilities of the witness and another 6% even believed that it depends **very little** on it. The remaining operators believed that the language abilities of the witness are important. 35% stated that it depends **very much** on the language abilities of the witness and 12% answered that it depends **much** on it.

How much does the construction of a facial composite depend on the memory abilities of the witness?

(n = 17) All of the operators stated that they believed that the memory abilities of the witnesses play an important role during the construction of the composite image. 76% **very much** believed this to be the case. The remaining responded that it depends **much** on the memory abilities of the witness.

How detailed are verbal facial descriptions provided by witnesses?

(n = 17) The majority of respondents (53%) reported that the verbal facial descriptions provided by witnesses are **moderately** detailed. 29% reported that they are **detailed**. 12% answered that they are **not detailed** at all. The remaining 6% believed that the descriptions are **very detailed**.

Witnesses have difficulties putting into words the description of the perpetrators' face.

(n = 16) 50% of the operators **agreed** with this statement. 25% **strongly agreed** with it. 19% of the operators **disagreed** with it and 6% were **undecided**.

Witnesses have difficulties understanding the instructions of the operator.

(n = 14) The vast majority of operators **disagreed** with this statement (86%). 7% even **strongly disagreed** with it. Only 7% **agreed** with it.

Witnesses have difficulties selecting individual features during the feature selection process.

(n = 15) The responses towards this statement varied considerably. 27% of the operators **agreed** with this statement, whereas another 27% **disagreed** with it. Further

27% were **undecided** about their answer. 13% **strongly agreed** with the statement and 7% **strongly disagreed** with it.

Witnesses have difficulties in constructing fine-grain changes.

(n = 16) 38% of the operators **agreed** with the statement that witnesses have difficulties in constructing fine-grain changes at the end of the composite construction process. Another 38% of the respondents were **undecided** about it. 13% **strongly agreed** with it. Only few operators **disagreed** (6%) or **strongly disagreed** (6%) with the statement.

Summary of main findings

The data of this survey study indicated that most operators find it easy to select facial features recalled by the witness during the composite construction phase. With regard to witness abilities, the majority of operators reported that the construction of the composite image depends only little on language abilities but very much on memory abilities of the witness. Most operators reported that the verbal facial descriptions provided by witnesses are only moderately detailed. With regard to difficulties witnesses might experience during the composite construction process, most operators agreed with the statement that witnesses might have difficulties with putting the description of the perpetrators' face into words. The vast majority of operators disagreed with the statement that witnesses might have difficulties understanding the instructions provided by the operator. Operators appeared unsure about whether witnesses might have difficulties in selecting individual features during the feature selection process, since the responses towards this statement varied a lot. Similarly, operators seemed to find it hard to decide whether witnesses might have difficulties in

constructing fine-grain changes to the face at the end of the composite construction process. Again, the answers to this statement were very diverse.

6.3.3 Experiences with and opinions about witnesses with LD

The following series of questions refers to the witness abilities of individuals with LD and any difficulties operators might face when engaging with LD witnesses during the composite construction process.

Have you ever generated a composite with a witness with LD?

(n = 17) 35 % of the operators have generated a composite image with the assistance of a witness with LD. The remaining 65% of the operators had not.

Are there any aspects of the facial composite system you are currently working with which are particular suitable for people with LD?

(n = 12) The response rate to this question was very low and the responses that were given fell into a variety of different subcategories; ‘experience with LD witnesses’, ‘type of composite system’ and ‘further specification of aspects particular suitable for LD witnesses’. Therefore, no percentages of responses are cited below, instead a detailed description of the raw data is provided.

Six operators, who had experience of constructing a composite with LD witnesses, responded to the question. Of those, four worked with E-FIT, one with ProFIT and EvoFIT and one operator with sketch and E-FIT. Two of the operators working with E-FIT answered that they did not know of any aspects of the E-FIT system which might be particular suitable for witnesses with LD. One E-FIT operator replied that the thumbnail option is particular suitable for LD witnesses. When asked to further specify the answer, the operator replied that the option to use the thumbnails

seems to work better for LD witnesses, since he/she can select the better feature option by comparing all options, rather than going through the different feature options sequentially and making final judgments after each presented feature option. The operator working with ProFIT and EvoFIT replied that there are no aspects of these programs which are particularly suitable for witnesses with LD. The operator using sketch and E-FIT replied that the freedom of choice on any feature and the opportunity to change it as many times as the witness requires were particularly suitable for people with LD. Unfortunately, it did not become clear from the answer to which composite system the two aspects referred to.

Six operators who had not experienced working with a witness with LD replied to the question. Five worked with E-FIT and one with ProFIT. The ProFIT operator responded that having one large facial image on the screen might be probably easier to view and to understand for an LD witness than when there are many smaller images on the screen. Only one E-FIT operator mentioned a suitable aspect of the E-FIT system for witnesses with LD, the fact that there is such a variety of features. Two E-FIT operators responded that there are no particular suitable aspects about the E-FIT system for witnesses with LD. And two further E-FIT operators replied that they do not know about any particular suitable aspects.

Are there any aspects of the facial composite system you are currently working with which are particularly unsuitable for people with LD?

(n = 11) Again, the raw numbers of responses to this question are described, meaning that no percentages of responses are cited.

Five operators with prior experience in generating a composite image with LD witnesses replied to this question. All of these 5 operators were working with E-FIT

and one also used sketch. Again, the replies from the E-FIT operators varied considerably. Three answered that there are no aspects of the E-FIT system which are particular unsuitable for LD witnesses. One E-FIT operator replied that he/she does not know about any aspects which are particular unsuitable for witnesses with LD. One E-FIT operator replied that most difficulties with LD witnesses seem to be memory based, and when presenting the changing facial options, these witnesses have a hard time retaining the original memory of the perpetrators' face. The operator working with sketch and E-FIT found it difficult to keep the LD witness focused on the feature being worked on. Unfortunately, it remains unclear whether this answer refers to sketch or the E-FIT system or whether the problem was experienced with both techniques.

The answers of the six operators with no previous experience with LD witnesses were less diverse. Five of these operators were working with E-FIT and one with ProFIT. Three of the E-FIT operators replied that they do not know about any aspects of the E-FIT system which might be particular unsuitable for witnesses with LD. One operator replied that the E-FIT system might be particular unsuitable for those LD witnesses with communication difficulties. Another operator responded that it might be difficult to make clear to the LD witness that they have to concentrate on one aspect of the face at a time. Furthermore, this operator mentioned that people with LD are 'peoples' pleaser' and that they might be therefore reluctant to make changes to the composite image. The ProFIT operator replied that the fact that there are no touch screen facilities available and that the composite images are not in colour and therefore appear flat might be particularly unsuitable for witnesses with LD.

How detailed are the verbal facial descriptions of witnesses with LD compared to witnesses without LD?

(n = 13) Most operators answered that the descriptions of witnesses with LD are **comparable in detail** to those provided by witnesses without LD (77%). The remaining 23% answered that the descriptions of LD witnesses are **less detailed** than those by non-LD witnesses.

(n = 6) Of those operators, who had generated a composite image with a witness with LD before, and can therefore base their answer on real experience, most reported that the descriptions provided by LD witnesses are **comparable in detail** to those provided by witnesses without LD (83%). 17% reported that the descriptions are **less detailed**.

(n = 7) A similar distribution of answers was obtained for the operators who had no experience with witnesses with LD. 71% answered that the descriptions by LD witnesses are **comparable in detail** and 29% assumed that the descriptions are **less detailed**.

Witnesses with LD have difficulties picturing the perpetrators' face in their mind.

(n = 13) Most of the operators were **undecided** about this statement (54%). 31% **disagreed** with it and 15% **agreed** with it.

(n = 6) Most operators, who had previously worked with a LD witness, **disagreed** with this statement (50%). 33% **agreed** with it and 17% were **undecided** about it.

(n = 7) Of those operators, who had never constructed a composite image with a witness with LD, 86% were **undecided** about whether witnesses with LD would

have difficulties picturing the perpetrators' face in their mind and 14% **disagreed** with this statement.

Witnesses with LD have difficulties putting into words the description of the perpetrator's face.

(n = 13) In general, most operators believed that witnesses with LD have some difficulties putting into words the description of the perpetrator's face. Of those, 15% **strongly believed** this to be the case and 38% **believed** it. The remaining operators were either **undecided** (31%) or **disagreed** with this statement (15%)

(n = 6) Of those operators, who were experienced with LD witnesses, 33% **strongly agreed** with the statement that LD witnesses have difficulties putting into words the description of the perpetrator's face. Another 33% **agreed** with it. 17% were **undecided** and 13% **disagreed** with it.

(n = 7) Of those operators who had no experience with LD witnesses, 43% **agreed** with the statement, another 43% were **undecided** and 14% **disagreed** with it.

Witnesses with LD have difficulties understanding the instructions provided by the operator.

(n = 13) Most operators were undecided regarding their answer (46%). However, a significant amount **disagreed** with it (23%). 8% even **strongly disagreed** with it. The remaining operators believed that LD witnesses have difficulties to understand the instructions by the operator. 15% **agreed** with the statement and 8% even **strongly agreed** with it.

(n = 6) Of those operators who were experienced with LD witnesses, 50% were **undecided**, 17% **disagreed**, and another 17% **strongly disagreed** with the statement. The remaining 17% **strongly agreed** with it.

(n = 7) Of those operators, who had not previously worked with a LD witness, 43% were **undecided**, 29% **agreed** with it, and 29% **disagreed** with it.

Witnesses with LD have difficulties selecting individual features during the feature selection phase.

(n = 13) 38% of the operators were **undecided** and 31% of the respondents answered that they **agree** with this statement. The other 31% **disagreed** with it.

(n = 6) The majority of operators with prior experiences with LD witnesses **disagreed** with the statement that they have difficulties selecting individual features during the feature selection phase (50%). 33% were **undecided** regarding their answer and 17% **agreed** with this statement.

(n = 7) 43% of the operators with no previous experience with LD witnesses **agreed** with the statement, another 43% were **undecided** about their answer and 14% **disagreed** with it.

Witnesses with LD have difficulties in constructing fine-grained changes of the face.

(n = 13) Most respondents were **undecided** (46%) regarding their answer to this statement. 23% **agreed** with it and another 23% **disagreed** with it. The remaining 8% **strongly agreed** with the statement.

(n = 6) Of those operators, who had created a composite image with a LD witness before, 33% were **undecided** about their answer and additional 33% **disagreed** with this statement. 17% **strongly agreed** with the statement and further 17% **agreed** with it.

(n = 7) Of those operators, who had no previous experience with witnesses with LD, 57% were undecided regarding whether LD witnesses may have difficulties

in constructing fine-grained changes to the composite. 29% **agreed** with the statement and 14% **disagreed** with it.

Are there any specific guidelines that can be referred to when generating a facial composite with the assistance of a witness with LD?

(n = 15) The majority of operators (53%) stated that they **do not know** about any specific guidelines. 40% of the respondents answered that **there are** specific guidelines available for generating a composite image with a LD witness. 7% of the operators answered that **there are no** guidelines available.

(n = 6) Of those operators, who had generated a facial composite with a LD witness before, 67% reported that they **do not know** about any specific guidelines. 17% reported that **there are no** guidelines available and further 17% answered that **there are** specific ones.

(n = 9) Of those respondents with no prior experience with LD witnesses, 56% reported that **there are** guidelines available, however, another significant amount answered that they **do not know** about any guidelines (44%).

Those operators, who answered that there are specific guidelines available for the composite construction with LD witnesses, were asked to specify those guidelines. The responses were very diverse. One operator stated “As with any composite, the witness can only do their best and the resultant composite is a representation of what the witness can recall. As long as everyone knows the witness has LD, the composite should carry as much weight as if the witness had no LD”. Operators referred to the *Facial Identification Guidance* document (NPIA, 2009), the *Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses and special*

measures document (CJS, 2007), the *Disability Discrimination Act* (1995 & 2005), the *Guidance of the management of police information* (MOPI) document (ACPO, 2006), the *Achieving Best Evidence document and the Youth and Criminal Evidence Act* 1999. One operator advised: “Keep the questions extremely simple. No multiple choices.” An operator with prior experiences with LD witnesses stated that he/she would only know about the appropriate adult scheme. Another operator with previous experiences with LD witnesses referred to the *Police and Criminal Evidence (PACE) Act* 1984 and quoted: “If a witness has the mental ability of a juvenile they must be treated as such and have an appropriate adult accompanying them.”

How frequently do you make reference to these guidelines?

(n = 4) Of those operators with prior experience with LD witnesses, **three** reported that they never make reference to these guidelines and one operator responded that he/she **sometimes** refers to them.

Summary of main findings

One third of the operators who completed the questionnaire had previous experience in generating a composite image with the assistance of a witness with LD. These operators were asked to base their answers on their own experiences. Those operators with no experience with LD witnesses were asked to respond to the questions with their best judgment. Overall, the majority of operators, those with and without previous experience with LD witnesses, answered that the verbal facial descriptions of witnesses with LD are comparable in detail to those provided by witnesses without LD. In general, most operators were undecided about whether witnesses with LD might experience difficulties in picturing the perpetrators' face in their mind during the composite construction process. Most operators, who previously had worked with

LD witnesses, disagreed with the statement, whereas the majority of operators, who had never constructed a composite image with a witness with LD, were undecided about it. In general, operators' responses to the statement that witnesses with LD have difficulties putting into words the description of the perpetrators' face were varied. This was true for operators with and without previous experiences with LD witnesses. Most operators were undecided whether witnesses with LD have difficulties understanding the instructions provided by the operator during the composite construction process. Operators provided varied responses to the statement whether witnesses with LD have difficulties selecting individual features during the feature selection phase. The majority of operators with experience with LD witnesses disagreed with the statement, while most operators with no experience with LD witnesses were undecided regarding their answer. A similar response pattern was obtained with regard to whether witnesses with LD have difficulties in constructing fine-grained changes on the face. Answers from experienced operators towards this statement were mixed, while the majority of inexperienced operators were undecided about the statement. A large proportion of operators stated that they do not know about any specific guidelines that can be referred to when generating a facial composite image with the assistance of a witness with LD. The majority of operators who had worked with a witness with LD before reported that they do not know about any specific guidelines. Most respondents without prior experience with LD witnesses reported that there are guidelines available. When further asked to specify these guidelines, the answers were very diverse; some operators were referring to official police documents, while others were giving recommendations based on their own knowledge or experience. The majority of experienced operators reported that they

never referred to such guidelines when they had constructed a facial composite with the assistance of a witness with LD.

6.4 Discussion

The question by question exploration of the data has highlighted several relevant aspects with regard to police operators' general practices and their experiences as well as attitudes towards witnesses with LD.

6.4.1 Current practices utilised by police operators during the composite construction

This survey study has revealed that UK police operators vary considerably in terms of their background and experiences; from the number of months of experiences as operators to the amount of facial composite images they have created during the past two years. This finding is in agreement with a similar survey of E-FIT operators conducted by Paine (2004). In Paine's study responses from operators ranged from 6 months to 17 years, with a mean length of operator's experience of 5.5 years. In a more recent survey study conducted by Brace, Pike and Turner (2008) comparable results were revealed. Operators facial composite construction experiences ranged from 1 to over 10 years. Very similar results were obtained during the present survey study.

The most widely used composite system in the UK appears to be E-FIT. The majority of operators during this survey study have stated that they either have experience with E-FIT or are currently working with it. This finding is in line with the results obtained by Brace et al. (2008). All of the operators participating in their survey were familiar with the E-FIT system. The present study revealed that the second mostly used facial composite systems appear to be EvoFIT and ProFIT.

Unfortunately, it was not possible to further examine whether this finding coincides with other statistics about frequently used composite systems in the UK, since an extensive literature search has revealed no available information.

All police operators have received some kind of training in the application of the facial composite software they are currently working with. Again this finding coincides with the results revealed by Brace et al. (2008), who found that almost 90% of their participants had received formal composite construction training and 10% training from another user. However, the present study did not identify a universal source of training. The majority of operators stated that they received training at the National Police Training Centre in Durham. Surprisingly, from the responses to the question regarding how long ago the facial composite training was, it became clear, that operators do not refresh their knowledge about facial composite systems and their applications on a regular basis, since answers to that question varied considerably. Nevertheless, it should be accredited that the UK provides some kind of professional training for police operators. A survey study by McQuiston-Surrett et al. (2006) for US police operators revealed that they do not have any standardised training available at all. Although, standardised training is available in the UK, no official legislations or guidelines exist on how often this training should be repeated or renewed to refresh existing knowledge and to remain up to date with current facial composite development and the latest empirical research findings.

With regard to how much the resulting quality of the facial composite depends on the abilities of the witness, most operators assumed that language abilities only play a minor role while memory abilities are of major importance. This is in contrast with the general procedure of how composite systems work. Most facial composite systems, such as E-FIT and ProFIT require a detailed description of the perpetrators

face at the beginning of the composite construction process (Davies & Valentine, 2006; Frowd et al., 2005b; see Chapter 4 for a detailed description of the composite construction process with computerised composite systems). Study 3 of this thesis revealed significant correlations between the amount of verbal information participants provided during the CIs about the target face and subsequent E-FIT composite likeness ratings scores. Thus, the completeness of the verbal facial description at the beginning of the composite construction determines how good the resultant composite will be. Interestingly, further verifying evidence comes from Study 5, which revealed that the composite construction performance of mLD witnesses can be considerably increased when a system is used that does not depend predominantly on language. The findings of the experimental studies in this thesis suggested that language and witness-operator communication certainly do play a crucial role during the composite construction, at least with systems currently used most frequently by UK police officers, such as E-FIT. According to most operators the verbal facial descriptions provided by witnesses are only moderately in detail. Finally, most operators believe that any arising difficulties witnesses might experience during the composite construction process are rather due to the witness having problems putting the description of the perpetrators' face into words, than due to difficulties understanding the instructions provided by the operator.

6.4.2 Operators' experiences with and attitudes towards witnesses with LD

This survey revealed that one third of the operators who have completed the questionnaire had previous experience in creating composite images with witnesses with LD. According to the operators, the facial descriptions provided by LD witnesses

are comparable in detail to those descriptions provided by witnesses without LD. This is in strong contrast to the empirical literature regarding the amount of information LD individuals' recall about an observed event. A consistently reported research finding is that the accounts of individuals with LD are considerably less detailed than accounts provided by individuals without LD (Agnew & Powell, 2004; Henry & Gudjonsson, 1999; Michel et al., 2000). With regard to facial recall, Study 2 and 3 of this PhD-project show that participants with mLD provide significantly less verbal information about a target face than participants without LD. It remains unclear why the respondents of this survey have the impression that verbal facial descriptions of LD witnesses are comparable in detail to those descriptions provided by witnesses without LD. When asking questions regarding the difficulties LD witnesses might experience during the facial composite construction process, operators without prior experience with LD witnesses provided diverse answers. In contrast, the majority of experienced operators were not aware of any difficulties LD witnesses might face during the composite construction. This finding is quite surprising, since one could assume that due to verbal as well as memory deficiencies witnesses with LD would experience more difficulties during the construction of facial composites than witnesses without LD.

Most operators seem to be unaware about specific guidelines they can refer to when generating a composite image with a LD witness. The majority of operators with prior experience with LD witnesses had no knowledge about any specific guidelines for the composite construction with witnesses with LD. When questioned about how often they would make reference to specific guidelines, most of the experienced operators responded with 'never'. This is rather concerning and highlights the importance of this PhD research project, since the obtained results

could have implications for police operators on how to obtain best evidence from witnesses with LD during the construction of facial composite images.

6.4.3 Practical problems operators' face during the composite construction with LD witnesses

One of the aims of this survey study was to identify practical problems police operators might experience during the construction of facial composites with LD witnesses. Therefore operators were asked whether any particular aspects of the facial composite system with which they are currently working might be particularly unsuitable for LD witnesses. Five operators with prior experience with LD witnesses responded to this question. Their responses are particularly valuable since they give insight into real-life practical problems operators face when creating a composite image with such witnesses. Four of these operators were working with E-FIT and one with sketch and E-FIT and their comments were limited to these systems. The answers of the E-FIT operators varied considerably. This might reflect the fact that people with LD do constitute a very heterogeneous group, and not every individual with LD displays the same cognitive capabilities. Three operators answered that there are no aspects of the E-FIT system which are particular unsuitable for witnesses with LD and one operator was unsure about whether there might some. Only one E-FIT operator mentioned a practical problem. According to this operator, most difficulties LD witnesses experience during the facial composite construction are due to their memory deficiencies. When presenting the changing facial feature options, LD witnesses seem to have a hard time retaining the actual memory of the perpetrators' face. This reflects the earlier reported tendency that most operators think that any difficulties arising during the composite construction with LD witnesses are due to

memory problems rather than problems with language or communication. The operator working with the sketch technique and the E-FIT system mentioned that it is very difficult to keep an LD witness focused on the facial feature being worked on. Thus, according to the operators, any arising practical problems during the construction of facial composites with the E-FIT system or the sketch technique originate from memory and attention deficits of LD witnesses rather than deficiencies in communication.

6.4.4 Limitations of the survey study

This survey study suffers from a number of limitations. The major limitation is one often associated with postal surveys. The actual response rate of UK police operators was very low. Furthermore, the author was not able to follow up incomplete questionnaires, since the names and addresses of the respondents remained anonymous. Questionnaires were initially sent by a second party and the agreement was that there would be no follow-up contacts. Due to the low response rate, it was obviously not possible to run any statistical tests on the data. Being restricted to a descriptive summary obviously limits the strength of any conclusions but it still has set a context for the experimental work presented later in the thesis. Although we were aware that a postal survey might result in low response rates, we still considered this sort of research tool as the only viable option for obtaining responses from numerous police operators from all over the UK. However, it is nevertheless disappointing that the response rate was so low.

A further short coming of this survey study was that the majority of respondents were E-FIT operators. Attempts were made to contact operators working with other systems but for whatever reason very few operators responded. Therefore the tentative findings are somewhat limited to the use of E-FIT.

Moreover, the study did not address how many witnesses with LD operators have had interviewed during their career. Retrospectively, answers to this question could have given insightful information and might have further emphasised the relevance of this research project.

Despite these shortcomings, the obtained data bares important implications for future research and provides a context to the current thesis. First, the findings of the survey study demonstrate that witnesses with LD are indeed placed into the situation where they have to create a facial composite image together with a police operator. Furthermore, the survey results show that there are no specific guidelines available for police operators on how to construct a composite image with such witnesses. Even operators with considerable composite construction experience, stated that they did not know whether there are any specific guidelines operators could refer to when working with LD witnesses. Providing standardised training courses on how to produce facial composites with LD witnesses appears to be therefore an important area of future policy and practice development, if accumulating evidence indeed shows that there are differences between LD and non-LD witnesses and their facial composite construction performance. Moreover, a high number of respondents chose the 'I don't know' answer option with regard to questions concerning witnesses with LD. This implies that a lot of police operators are unsure about how to treat witnesses with LD in an appropriate manner, which highlights even more the need of conducting research regarding LD witnesses' composite construction abilities and measures which might facilitate their performance.

Chapter 7

Study 2: Face recognition and description abilities in people with mLD of unfamiliar faces

This chapter describes the first experimental study, which compares the ability of people with mLD and control participants to recognise and describe unfamiliar faces. The study consists of two experiments. Experiment 1 includes three old/new face recognition tasks and Experiment 2 consists of two face description tasks. The aim of the study was to gather insight into basic face recognition and description abilities of people with mLD.

7.1 Introduction

Humans have a remarkable ability to encode new faces. One of the first studies demonstrating the outstanding ability of humans to recognise unfamiliar faces was conducted by Yin (1969). Participants viewed pictures of unfamiliar faces for 3 to 5 seconds and subsequently had to engage in an old/new face recognition task. Participants performed at about a 90% success rate for a series of pictures ranging from 8 to 144 photographs. Since then, more than thousands of studies have been carried out investigating face recognition and the factors influencing it (Ellies, 1975; see Chapter 4 for a review of applied research in face recognition). In contrast, only a few studies have been conducted investigating face recognition and description abilities of people with LD. One of the first studies that examined face recognition skills in individuals with LD was carried out by McCartney (1987). He investigated memory for faces in teenagers with and without LD. Participants were tested either immediately, one day later, or one week later, using a forced choice-recognition test.

Overall, LD participants performed at a lower level than the non-LD group. However, there was no differential memory loss between the two participant groups. A six months follow-up test revealed that although the LD group once again performed poorer, overall their memory loss was comparable to that of the non-LD group. On the basis of these results McCartney (1987) concluded that there are no long-term memory differences between individuals with and without LD for face stimuli. In another study conducted by Dobson and Rust (1994) forgetting rates of participants with and without LD on two different memory tasks were compared; memory for objects and memory for faces. It was found that, in a recognition task, participants with LD had more difficulties learning the object stimuli to a criterion of 100% than non-LD individuals did. However, no group differences were obtained for the number of learning trials required for faces. Moreover, both groups required significantly fewer learning trials to learn the face than the object stimuli. All participants remembered significantly more faces than objects after a time delay of 1 week, 1 month and 2 months. There were no significant differences between the LD and the non-LD groups in memory for faces on any of the re-test trials. These results suggest that different processes are used during the recognition of faces versus the recognition of objects. The notion that face recognition draws on different cognitive processes than object recognition is not a new argument and it forms an important area in face recognition research (e.g. Diamond & Carey, 1986; Johnson & Morton, 1991; Scapinello & Yarmey, 1970; Yin, 1996). However, Dobson and Rust (1994) were one of the first who explored face recognition abilities in individuals with LD and demonstrated that this face recognition mechanism is distinct and well developed in both people with and without LD.

Research in the eyewitness domain looking particularly at face description abilities in people with LD is rare, although good face description skills might play a crucial role in the successful completion of criminal investigations. For instance, these abilities form a prerequisite for the accurate construction of facial composite images. Milne, Clare and Bull (1999) examined the type of information people with and without LD were providing about an observed event, which depicted an accident where a boy got knocked down by a car. They differentiated between person, object and action details. In general, both participant groups reported fewer and less accurate person details than object and action details. However, during this study person details were not further subdivided into information about the person's face. Therefore, it remains unclear whether participants mentioned any facial information at all about the person depicted in the video and how potentially useful this information might have been in a forensic context.

To guarantee that individuals with LD are treated in a fair and reasonable way by the criminal justice system and have access to the same procedures as witnesses without LD, it is important that more research is conducted into their ability to engage in tasks that are part of the investigation process, such as recognising and describing faces. The current series of experiments intends to shed more light on basic face recognition and description skills of witnesses with mLD.

7.1.1 Research aims of Study 2

The aim of this study was to gather insight into basic face recognition and description abilities of people with mLD and in particular:

- 1) To investigate the ability to recognise previously seen unfamiliar faces and to compare this ability to that of people without LD.

- 2) To establish the ability of people with mLD to describe unfamiliar faces and to compare the language and terms used with those used by witnesses without LD.
- 3) To investigate the effect of questioning type (free vs. cued recall) on the quantity and quality of the facial descriptions. It was hypothesised that regardless of the group (mLD vs. controls), participants would provide more information during the cued recall than during the free recall condition. It is also hypothesised that participants would give more accurate information during the free recall compared to the cued recall.
- 4) To explore the effect of memory performance on participant's descriptions (photo vs. memory condition). The hypothesis is that regardless of the group (mLD vs. controls), participants will perform more accurately and provide more detailed facial descriptions during the photo condition than during the memory condition. With regard to group, it was hypothesised that participants with mLD will perform as accurately as participants without LD during both description modes (photo vs. memory), however their descriptions will be less complete.

7.2 Experiment 1 Face recognition skills in people with mLD

This experiment investigates the ability of people with mLD to recognise previously seen unfamiliar faces and compares their ability to that of individuals without LD.

7.2.1 Method

Participants

Sixty participants took part in this study. Thirty were people with mLD who were recruited from the Kemback ARC in Dundee (21-58 years; $M = 39$ yrs; $SD = 10.71$; WASI: FSIQ-4 score: $M = 58.48$, $SD = 5.28$; WASI: verbal score: $M = 57.28$, $SD = 4.59$; WASI: performance score: $M = 64.97$, $SD = 6.53$) and 30 were students and members of staff from the University of Abertay Dundee (19-54 years; $M = 29$ years; $SD = 8.89$). All individuals with mLD had a WASI: FSIQ-4 score between 50 and 70 and therefore lay in the classification range of mLD as utilised by the WHO.

Psychometric tests

Verbal as well as non-verbal performance and general intellectual functioning of the mLD group was assessed with the Wechsler Abbreviated Scale of Intelligence (WASI). The WASI is a short form of the WAIS and consists of four subtests: the Vocabulary and Similarity subtests assess verbal abilities and the Matrix reasoning and Block design subtests measure visual-motor and coordination skills. Together, the subtests provide an estimate of general intellectual ability and can be administered in approximately 45 minutes. The experimenter received training in the administration of the WASI from an experienced clinician and can be therefore regarded as competent to administer this test.

Design

A between-subjects design was employed, including the between-subject factor group (mLD vs. control). The dependent variable was the accuracy of recognition performance (accuracy of old/new judgments).

Materials

For practice task 1, full-face photographs of Caucasian females were used as stimuli. Practice task 2 included Mr Men characters as stimuli. During the main recognition task, full-face photographs of Caucasian females served again as stimuli. All facial stimuli were photographed without spectacles or other distinguishing marks and the facial expressions were all neutral. Each picture was 564 x 765 pixels in size. Faces were shown from the front. All facial stimuli were unfamiliar to the participants (see Figure 7.1, 7.2 and 7.3 for pictures of the stimuli utilised).

Procedure

The experimental procedure was the same for both groups of participants. Participants took part individually. All stimuli were presented on a monitor of a Toshiba laptop running Superlab software. The size of the laptop was 36.2 cm x 26.8 cm x 3.9 cm with a screen resolution of 1280 x 800. The face recognition experiment consisted of two practice tasks and the main recognition task. Each task included two phases, a learning phase, during which participants had to memorise the presented stimuli, and a test phase, during which participants had to recognise the stimuli seen previously during the learning phase. During the learning phase, stimuli were presented for 10 seconds. During the test phase, stimuli were on the screen for as long as participants needed to make their old/new judgments. In all the tasks, the experimenter provided all the instructions to the participants orally. The instructions were kept simple and participants were asked to repeat back what was requested from them to ensure that they understood the instructions completely. Furthermore, they were encouraged to ask questions if anything was unclear at any time. All stimuli were presented to

participants in a random and sequential order and all answers of the participants were entered manually by the experimenter and recorded via the Superlab software.

Practice task 1: During the first practice task, participants were presented with one face during the learning phase and subsequently with the same face and an entirely new face during the test phase (see Figure 7.1 for an example of practice task 1). Both faces appeared as a target the same number of times during the course of the experiment. In the learning phase, participants were asked to remember the presented face. In the subsequent test phase, participants were asked to indicate whether they had seen this face before or whether this was an entirely new face. Hence participants engaged in a two alternative forced choice task. The purpose of this was to get participants used to the general procedure and the instructions provided by the experimenter.

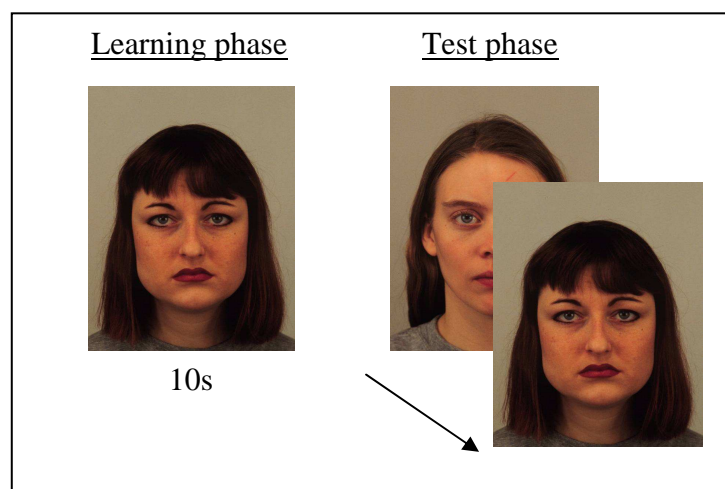


Figure 7.1. *Practice task 1. During the learning phase, one female face is presented. In the subsequent test phase, two female faces are presented, of which one is old and one is new.*

Practice task 2: The general procedure during the second practice task was the same as during the first one. The purpose of this second practice task was to increase the demands of the practice by increasing the number of stimuli. Mr Men characters were used as stimuli instead of faces to decrease any potential interference effects during

the main recognition task. Participants were presented with three Mr Men characters during the learning phase and six Mr Men characters, of which three were old and three were new, during the test phase (see Figure 7.2 for an example of practice task 2). An inter-stimulus interval (ISI) of one second was applied. Between trials, a brief cue was presented, to direct participant's attention to the centre of the screen.

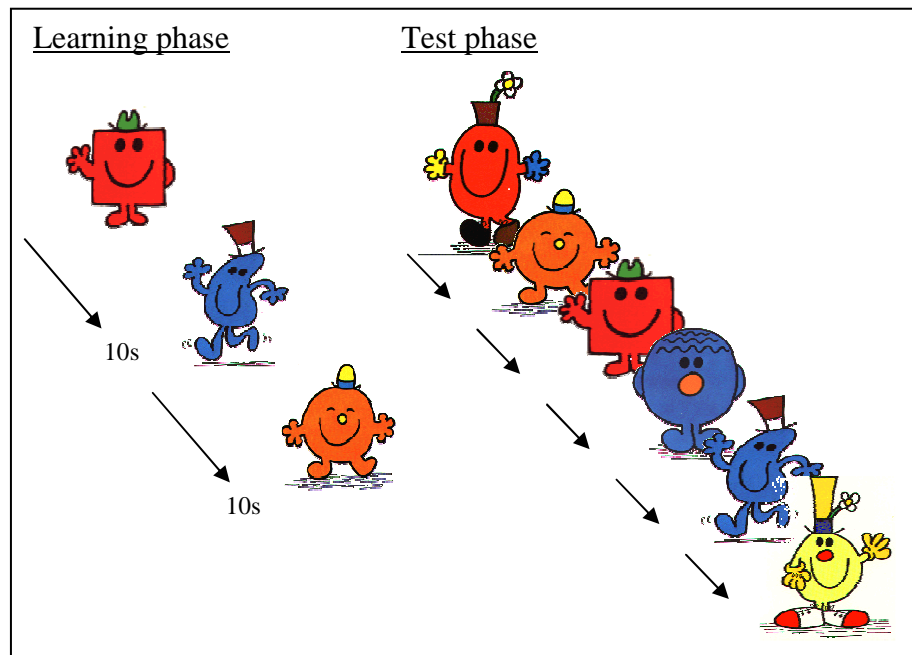


Figure 7.2. *Practice task 2. During the learning phase, three Mr Men characters are presented. In the test phase, six Mr Men characters are presented, of which three are old and three are new.*

Main recognition task: In the main recognition task, participants were assigned to one of two conditions. Each condition included the presentation of a different set of five faces during the learning phase. As during previous tasks, participants were asked to remember the presented faces (see Figure 7.3 for an example of the main recognition task). An ISI of one second was applied and a brief cue was presented between stimuli to direct the attention of the participants towards the middle of the monitor screen. Directly, thereafter, the test phase took place. Ten faces (five faces previously presented during the learning phase and five entirely new ones) were

presented one at a time and participants were asked to indicate whether they had seen the presented face before or whether it was a new face.

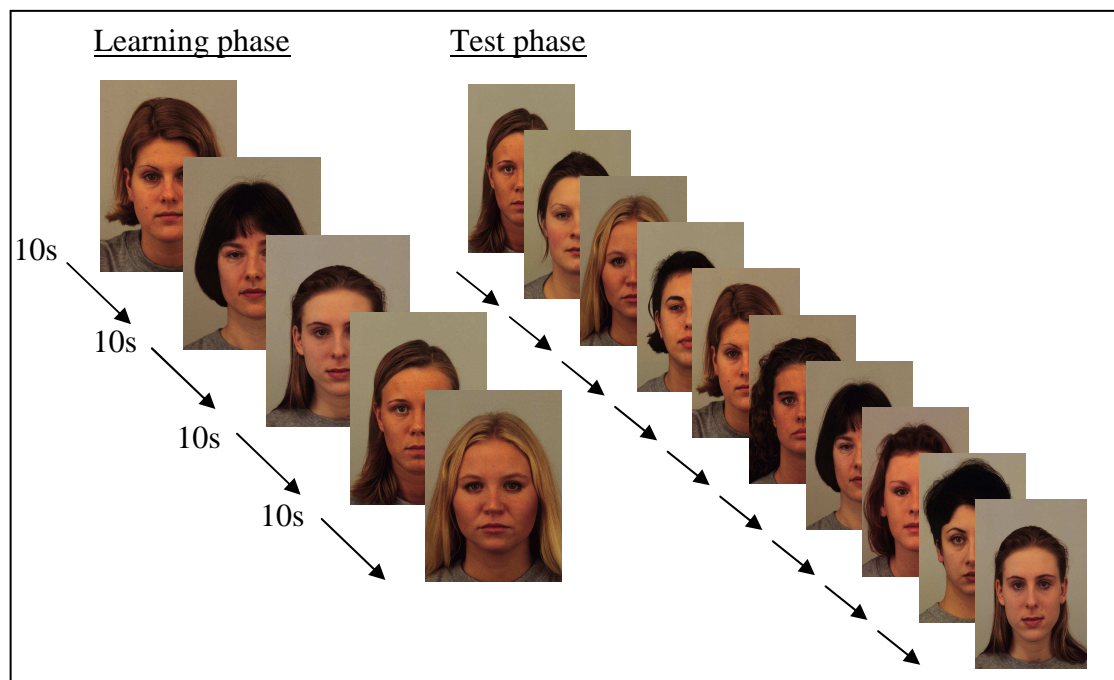


Figure 7.3. *Main recognition task. During the learning phase, five faces are presented. In the test phase, 10 faces are presented, of which five are old and five are new.*

Scoring

For all tasks the total amount of correct and erroneous responses were calculated for each participant.

7.2.2 Results

Research Questions

Data analysis focused on the following research questions: First, is the performance of people with mLD comparable to that of individuals without LD? Second, are individuals with mLD able to recognise a previously seen unfamiliar face? If that is the case, their performance on the three recognition tasks should exceed performance by chance alone. An alpha level of .05 was used for all statistical tests.

Practice task 1

To analyse the data obtained during the first practice task it was coded whether participants' performance was erroneous or error free. Nineteen out of 30 participants with mLD performed error free on this task, whereas all control participants performed error free. A Chi-square test was conducted to examine whether there was a significant association between the performance at the task and whether the participant had mLD or not. The test revealed a significant association between task performance and whether or not participants had mLD $\chi^2(1, N = 60) = 13.47, p < .001$.

To further investigate whether participants with mLD performed better than would be expected by chance alone, a binomial test was carried out. The test revealed that the performance of mLD participants was not significantly better than would be expected by chance alone, $z = 1.46, p = 0.20$.

Practice task 2

During practice task 2, participants with mLD had on average 4.40 ($SD = 1.30$) items correct out of 6 (hits and correct rejections were collapsed), whereas control participants had 5.69 items correct ($SD = 0.54$). To investigate whether there was a significant difference in performance between the mLD group and the control group on this task, an independent t-test was conducted. The independent variable was group (mLD vs. control) and the dependent variable was recognition accuracy (a score out of six). The Levene's test of the t-test was significant ($p < .001$), indicating that the variances in the two populations were not equally distributed. Further elaboration of the data showed that the data was also not normally distributed; the Kolmogorov-Smirnov test (K-S test) was significant ($p < .001$). Therefore, a nonparametric

equivalent to the independent t-test was applied; the Mann-Whitney test. It was found that people with mLD did differ significantly in their performance during practice task 2 from control participants, in that they performed significantly poorer (mLD: $Mdn = 5$, control: $Mdn = 6$), $U = 185.50$, $p < .001$, $r = -.53$.

To investigate whether the performance of individuals with mLD was better than chance normally the critical region of the binomial distribution would be calculated. To use the normal distribution to determine critical values, both pn and qn must be at least 10 (see Gravetter & Wallnau, 2007, page 632). However, in this case the binomial distribution had a mean of $pn = (1/2)(6) = 3$. Therefore, it was not possible to calculate the critical region, which would indicate whether an individual was scoring significantly different from chance. Instead, the critical region for the whole group of participants was determined to examine whether individuals with mLD as a group were scoring significantly different from chance. The binomial distribution had a mean of $pn = (1/2)(30 \times 6) = 90$ and a standard deviation of $\sqrt{npq} = \sqrt{180(1/2)(1/2)} = 6.71$. To be significantly different from chance, the score must be above (or below) the mean by at least $1.96(6.71) = 13.15$. Thus, with a mean of 90, the group would need to score above 103.15 ($90 + 13.15$) or below 76.85 ($90 - 13.15$) to be significantly different from chance. The group score on practice task 2 was 132, which is significantly above chance performance. Thus, although individuals with mLD performed significantly poorer than their non-LD counterparts, their performance as a group was significantly better than would be expected by chance alone. This suggests that the mLD group understood the task. On this basis it was deemed appropriate to move to the main experiment.

Main recognition task

In the main recognition task, mLD participants had on average 7 ($SD = 2$) items recognised correctly out of ten and participants without LD 9.77 ($SD = 0.50$) (again hits and correct rejections were collapsed). To examine whether there was a significant difference in performance between the two experimental groups on the main recognition task an independent t-test was carried out. Again the data were not normally distributed and the variances across groups were not equal; the K-S test and the Levene's test were significant (both $ps < .001$). Therefore, the Mann-Whitney test was utilised. The results showed that participants with mLD performed significantly poorer on the main recognition task than individuals without LD (mLD: $Mdn = 7$, control: $Mdn = 10$), $U = 91.50$, $p < .001$, $r = -.72$.

To further investigate why people with mLD performed poorer than their non-LD counterparts, the overall amount of hits (responding 'old' to an old item) and the number of correct rejections (responding 'new' to a new item) were analysed separately. Since previous research has found that people with mLD often show a high tendency of acquiescence (responding to questions affirmatively) (Gudjonsson, 1990), it was expected that they would show a high proportion of hits but only few correct rejections. The total amount of hits and correct rejections was compared for each group separately with a Wilcoxon signed-ranks test. The test revealed that the mLD and non-LD groups did not differ significantly in their total amounts of hits and correct rejections. Individuals with mLD had on average 3.97 ($SD = 1.59$, $Mdn = 5$) hits and 3.03 ($SD = 2.14$, $Mdn = 4$) correct rejections, $T = 102.00$, $p = 0.10$, $r = -0.29$. Individuals without LD had on average 4.90 hits ($SD = 0.31$, $Mdn = 5$) and 4.87 ($SD = 0.35$, $Mdn = 5$) correct rejections, $T = 6$, $p = 0.65$, $r = -0.08$. Thus, a higher tendency

of acquiescence was in this case not responsible for the poorer performance of the mLD group.

To examine whether the group performance of individuals with mLD was better than chance, the critical region of the binomial distribution was determined. The binomial distribution had a mean of $pn = (1/2)(30 \times 10) = 150$ and a standard deviation of $\sqrt{npq} = \sqrt{300(1/2)(1/2)} = 8.66$. To be significantly different from chance, the groups score must be above (or below) the mean by at least $1.96(8.66) = 16.97$. Thus, with a mean of 180, the group would need to score above 166.97 ($150 + 16.97$) or below 133.03 ($150 - 16.97$) to be significantly different from chance. The group's mean score was 210. So, even though, mLD participants performed significantly poorer than participants without LD, their group performance was significantly above chance performance.

Correlations between IQ and performance

Correlations between the verbal-, performance-, and full WASI score (FSIQ-4 score) and the performances during the different recognition tasks were calculated for participants with mLD. For practice task 1 no significant correlations between performance and individuals IQ were obtained (all $ps > .05$). For practice task 2, only one significant correlation was revealed between the verbal WASI score and mLD individual's performance, $r = .428$, $p = .021$. Thus, mLD individuals with a higher verbal WASI score obtained more correct responses during this task than mLD individuals with lower verbal WASI scores. For the main recognition task, all WASI scores were significantly correlated with recognition accuracy (verbal WASI score: $r = .404$, $p = .030$; performance WASI score: $r = .410$, $p = .027$; full WASI score: $r = .435$, $p = .018$). Thus, mLD participants with higher WASI scores performed better during the main recognition task than mLD participants with a lower score.

Correlations between performance on practice task and the main recognition task

Significant correlations were obtained between mLD and non-LD individuals' performance on practice task 2 and the main recognition task (mLD: $r = .662$, $p < .001$; controls: $r = .494$, $p = .006$). Thus, people performing better on the practice task performed also superior during the main recognition task.

Summary

Participants with mLD performed significantly poorer than control participants on practice task 1. Furthermore, at an individual level, a significant amount of mLD participants performed merely at chance level. Therefore, it was necessary to include a second practice task, to train and prepare participants sufficiently for the main recognition task. The results of practice task 2 showed that participants with mLD as a group were performing better than by chance, though it should still be noted that a substantial number was scoring around what would be expected by chance. The group performance indicates that mLD participants understood the questions asked by the experimenter and managed to perform the task in line with the instructions. However, despite performing above chance level they performed less well than control participants. During the main recognition task, participants with mLD performed significantly poorer than their non-LD counterparts. However, their group performance was again above chance level, indicating that they were able to understand the task and to accurately recognise previously seen unfamiliar faces. See Table 7.1 for an overview of the data.

Table 7.1. Overview of findings obtained during Experiment 1. Count of error free performers on practice task 1 and mean scores for mLD- and control participants (+SD) on practice task 2 and the main recognition task.

Group	Practice task 1	Practice task 2		Main recognition task	
	Count of error free performers ^a	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
mLDs	19	4.40	5.73	7	2
Controls	30	5.69	.54	9.77	.50

^a n = 30 for each group.

7.3 Experiment 2 Face description skills in people with mLD

During this experiment, the ability of people with mLD to describe unfamiliar faces was explored and compared to the performance of individuals without LD.

7.3.1 Phase 1 Describing unfamiliar faces

Method

The same participants took part as in the first experiment.

Design: A 2 (group: mLD vs. control) x 2 (description mode: memory vs. photo) x 2 (recall condition: free vs. cued) mixed design was used, including one between-subject factor (group) and two within-subject factors (description mode and recall condition). The dependent variable was the quantity and quality of the facial information provided by the participants.

Material: Stimuli were full-face photographs of Caucasian males. There was a total of six faces. All faces were photographed without spectacles or any other distinguishing marks, the facial expressions were neutral and all faces were shown from the front. Each picture was 216 x 295 pixels in size. Examples of the face stimuli

are depicted in Appendix 2. The stimuli derived originally from a database of images of police trainees created by the UK Home Office Police Information Technology Organisation (PITO). The faces were used as stimuli in previous peer reviewed research by Bruce, Henderson, Greenwood, Hancock, Burton and Miller (1999).

Procedure: Participants were asked to describe a picture of a face, which was presented on a monitor of a Toshiba laptop. The size of the laptop was 36.2 cm x 26.8 cm x 3.9 cm with a screen resolution of 1280 x 800. Each participant engaged in two different description modes (memory vs. photo). During the memory condition, the presented face disappeared after 10 seconds and participants were required to describe the face from their memory alone. In the photo condition, the presented face was visible all the time and participants were encouraged to look as often and as long at the face as they needed to, while providing the description. The memory condition always preceded the photo condition to avoid memory interference. During both description modes, participants engaged in two recall conditions: a free recall and a cued recall. During the free recall, participants were asked to report everything they could about the presented face. During the cued recall, participants were asked more specific questions about the different individual facial features (e.g., “What did the hair of the face look like?” or “What did the nose of the face look like?”). The free recall always preceded the cued recall. During each description mode, participants were presented with a different face (out of six faces). Each participant viewed two faces in total. The selection of these two faces was different for each participant. Participants were asked to indicate when they had finished the verbal description. No time limits were set for the task. The verbal responses of each participant were tape recorded with an Olympus VN-3100PC Digital Voice Recorder. At the end of the task participants were fully debriefed and thanked for their participation.

Scoring:

Quantity of facial information. To investigate the quantity of facial information all verbal descriptions were transcribed and the overall amount of facial information mentioned during the free recall and the cued recall was counted for both participant groups.

Quality of facial information. To examine the quality of the facial information for each target face, a checklist was created, containing the accurate descriptions of the individual facial features. The checklist was based on a pilot study, during which an independent sample of participants (16 participants overall) decided which descriptions were most appropriate for the different facial features of each of the six target faces. The descriptions were taken from the Aberdeen Index which is used in the E-FIT composite system. The labels selected most often by the independent sample were regarded as the accurate ones and included in the checklist. A similar approach was used by Schooler and Engstler-Schooler (1990) to determine the accuracy of verbal facial descriptions provided by participants in their study. Thus, to determine the quality of the facial descriptions, the information included in the checklist was compared with the information provided by the participants during the actual experiment.

Inter-coder reliability: To determine the inter-coder reliability, two coders coded the facial descriptions provided by the participants independently. The correlational analysis showed a significant level of agreement between the two coders based on a random sample of eight transcripts. See Table 7.2 for exact levels of agreement between the two coders.

Table 7.2. *Correlations obtained between the two coders during the different description modes and recall conditions based on 8 transcripts (Spearman's rho correlation coefficients and p values are provided).*

	Description mode							
	Memory				Photo			
	Free recall		Cued recall		Free recall		Cued recall	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>P</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Quantity of facial information	.98	<.001	.98	<.001	.94	.001	.98	<.001
Quality of facial information	.96	<.001	.97	<.001	.89	.003	.96	<.001

Results

Research Questions: The analysis of data examined the following research questions: First, when describing an unfamiliar face, do individuals with mLD differ to those without LD with regard to the quantity and quality of facial information they provide? Second, do mLD participants perform in a similar manner as non-mLD participants during the different description modes and recall conditions? An alpha level of .05 was used for all statistical tests.

Quantity of facial information: The average amount of facial information provided by the two participant groups during the different description modes and recall conditions was calculated and the data are depicted in Table 7.3.

Table 7.3. *Mean number of facial information (+SD) provided by mLD and control participants during the different description modes (photo vs. memory) and recall conditions (free vs. cued).*

	Description mode							
	Memory				Photo			
	Free recall		Cued recall		Free recall		Cued recall	
	<u><i>M</i></u>	<u><i>SD</i></u>	<u><i>M</i></u>	<u><i>SD</i></u>	<u><i>M</i></u>	<u><i>SD</i></u>	<u><i>M</i></u>	<u><i>SD</i></u>
mLDs	1.96	2.57	4.57	3.68	3.50	3.43	5.64	2.98
Controls	7.13	3.72	12.87	4.55	11.97	3.86	14.72	4.35

To investigate whether people with mLD provide less facial information during their descriptions than control participants a 2 x 2 x 2 mixed ANOVA was conducted. The two within subject-factors were description mode (photo vs. memory) and recall (free vs. cued). The between-subject factor was group (mLD vs. control). The Levene's test was significant for two of the four dependent variables (photo & cued recall: $p = .037$; memory & free recall: $p = .027$), indicating that the variances were not distributed equally across groups and hence violated one of the assumptions of the parametric test. The data was log transformed (ln) to correct for this violation. Despite this, the Levene's test remained significant (photo & cued recall: $p = .033$; memory & free recall: $p = .030$). Therefore, the data were analysed with separate non-parametric Mann-Whitney tests. The findings revealed that during all description modes (photo & memory) and recall conditions (free recall & cued recall), mLD participants mentioned significantly less facial information than members of the control group (memory & free recall: $U = 89.50$, $p < .001$, $r = -.68$; memory & cued recall: $U = 67.00$, $p < .001$, $r = -.72$; photo & free recall: $U = 45.50$, $p < .001$, $r = -.77$; photo & cued recall: $U = 27.00$, $p < .001$, $r = -.80$) (mLD: memory & free recall: $Mdn = 1$, memory & cued recall: $Mdn = 4$, photo & free recall: $Mdn = 3$, photo & cued recall: $Mdn = 5.5$; controls: memory & free recall: $Mdn = 7$, memory & cued recall: $Mdn = 13$, photo & free recall: $Mdn = 11$, photo & cued recall: $Mdn = 14$) (see Figure 7.4).

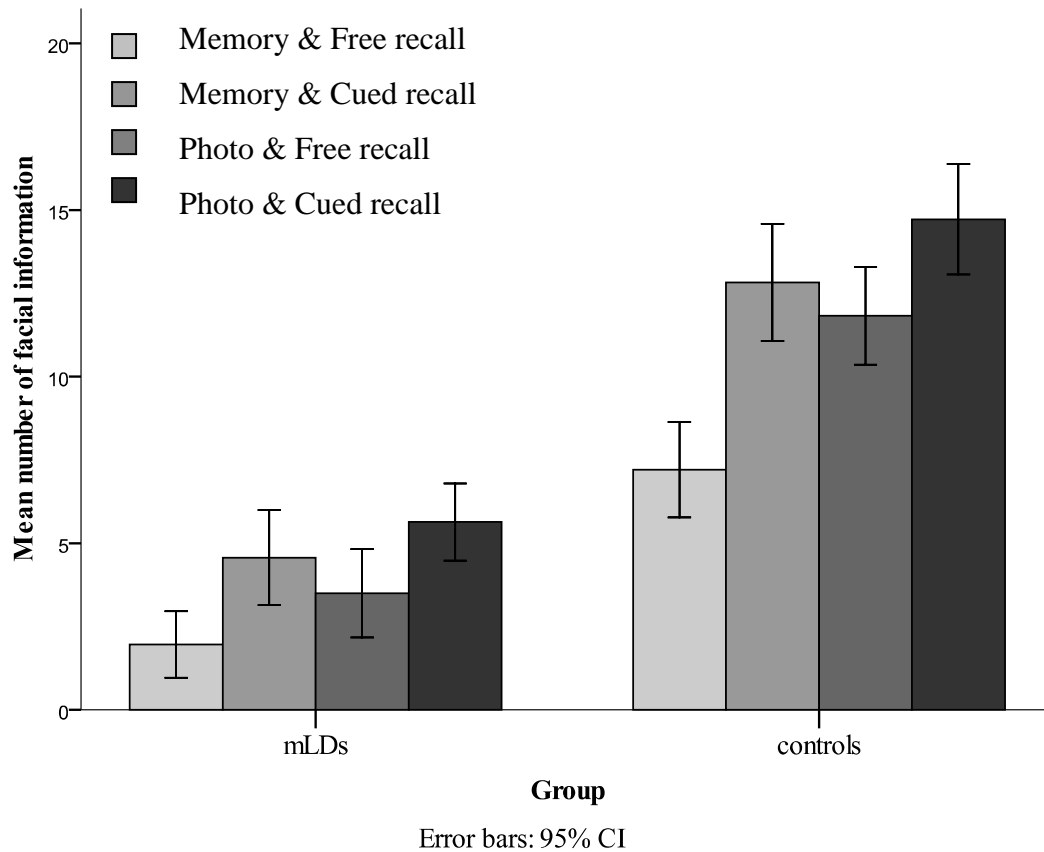


Figure 7.4. Mean number of facial information (+SE) provided by mLD- and control participants during the different description modes (photo vs. memory) and recall conditions (free vs. cued).

To further examine whether each of the two participant groups performed differently during the two description modes and the two recalls, the data were collapsed for description mode and recall condition and analysed separately for each participant group (mLDs & controls) with a Wilcoxon signed-ranks test. For the two different description modes it was found that both participant groups mentioned significantly more information during the photo (mLD: $Mdn = 8.5$, control: $Mdn = 26$) than during the memory condition (mLD: $Mdn = 6$, control: $Mdn = 19$) (mLD: $T = 34.50$, $p < .001$, $r = -.68$; control: $T = 25.50$, $p < .001$, $r = -.79$). For the two recall conditions, it was obtained that both participant groups provided significantly more facial information during the cued recall (mLD: $Mdn = 9$, controls: $Mdn = 27$) than

during the free recall (mLD: $Mdn = 4$, controls: $Mdn = 19$) (mLDs: $T = 3.50$, $p < .001$, $r = -.83$; controls: $T = 2$, $p < .001$, $r = -.88$) (see Table 7.4).

Table 7.4. Mean number of facial information (+SD) provided by mLD and control participants during the different description modes (photo vs. memory) and recall conditions (free vs. cued). The data are collapsed across conditions.

	Description mode				Recall			
	Memory		Photo		Free recall		Cued recall	
Group	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
mLDs	6.54	5.73	9.14	5.88	5.46	5.71	10.21	6.01
Controls	20.00	7.11	26.55	7.50	19.10	6.27	27.55	8.24

Quality of facial information: To examine the quality of facial information provided by the two participant groups during the different description modes and recall conditions accuracy rates were calculated and expressed here as percentages. This was done by dividing the amount of accurate facial information by the overall provided amount of information and multiplying the result with 100. The data are shown in Table 7.5.

Table 7.5. Mean percentages of accurate facial information (+SD) provided by mLD and control participants during the different description modes (photo vs. memory) and recall conditions (free vs. cued).

	Description mode							
	Memory				Photo			
	Free recall		Cued recall		Free recall		Cued recall	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
mLDs	54.94	41.34	50.18	30.14	54.47	30.35	55.02	23.13
Controls	71.74	17.58	69.95	13.93	75.83	14.05	70.80	13.14

A 2 x 2 x 2 mixed ANOVA was carried out to examine whether people with mLD mentioned less accurate facial information during their descriptions than control participants. The Levene's test was significant (all $ps < .05$), implying that the

variances across the different groups were not equally distributed. To correct for this violation the data were log transformed (ln). However, the Levene's test remained significant (all $ps < .005$). Therefore, the Mann-Whitney test was utilised to analyse the gathered data. It was found that during the majority of description modes and recall conditions, mLD participants mentioned significantly less accurate information than control participants (memory & cued recall: $U = 202.00$, $p = .002$, $r = -.41$; photo & free recall: $U = 167.50$, $p = .005$, $r = -.38$; photo & cued recall: $U = 210.50$, $p = .003$, $r = -.40$) (mLDs: memory & cued recall: $Mdn = 47\%$, photo & free recall: $Mdn = 50\%$, photo & cued recall: $Mdn = 53\%$; controls: memory & cued recall: $Mdn = 73\%$, photo & free recall: $Mdn = 77\%$, photo & cued recall: $Mdn = 71\%$ (see Figure 7.5). No significant difference between groups was obtained for the amount of accurate information provided during the memory and free recall condition (mLDs: $Mdn = 45\%$, controls: $Mdn = 74\%$) ($U = 223.50$, $p = .320$, $r = -.13$) (see Table 7.5).

To further explore whether there were any differences in the amount of accurate facial information participants provided during the two description modes and the two recall conditions, the data were collapsed for description mode and recall condition and analysed separately for each participant group (mLDs & controls) with a Wilcoxon signed-ranks test. mLD participants and control participants did not differ significantly in the amount of appropriate facial information provided during the different description modes (mLDs: memory: $Mdn = 45\%$, photo: $Mdn = 55\%$; controls: memory: $Mdn = 73\%$, photo: $Mdn = 75\%$) (mLDs: $T = 132$, $p = .412$, $r = -.16$; controls: $T = 194.00$, $p = .611$, $r = -.09$). A similar finding was obtained with regard to the different recall conditions (mLDs: free recall: $Mdn = 52\%$, cued recall: $Mdn = 50\%$; controls: free recall: $Mdn = 75\%$, cued recall: $Mdn = 70\%$) (mLDs: $T = 123.50$, $p = .66$, $r = -.08$; controls: $T = 117.00$, $p = .05$, $r = -.36$). However, with

regard to description mode, the trends point into the direction that both groups mentioned more accurate information during the photo condition than during the memory condition. Regarding recall condition, both groups provided more accurate information during the free recall condition compared to the cued recall (See Figure 7.5).

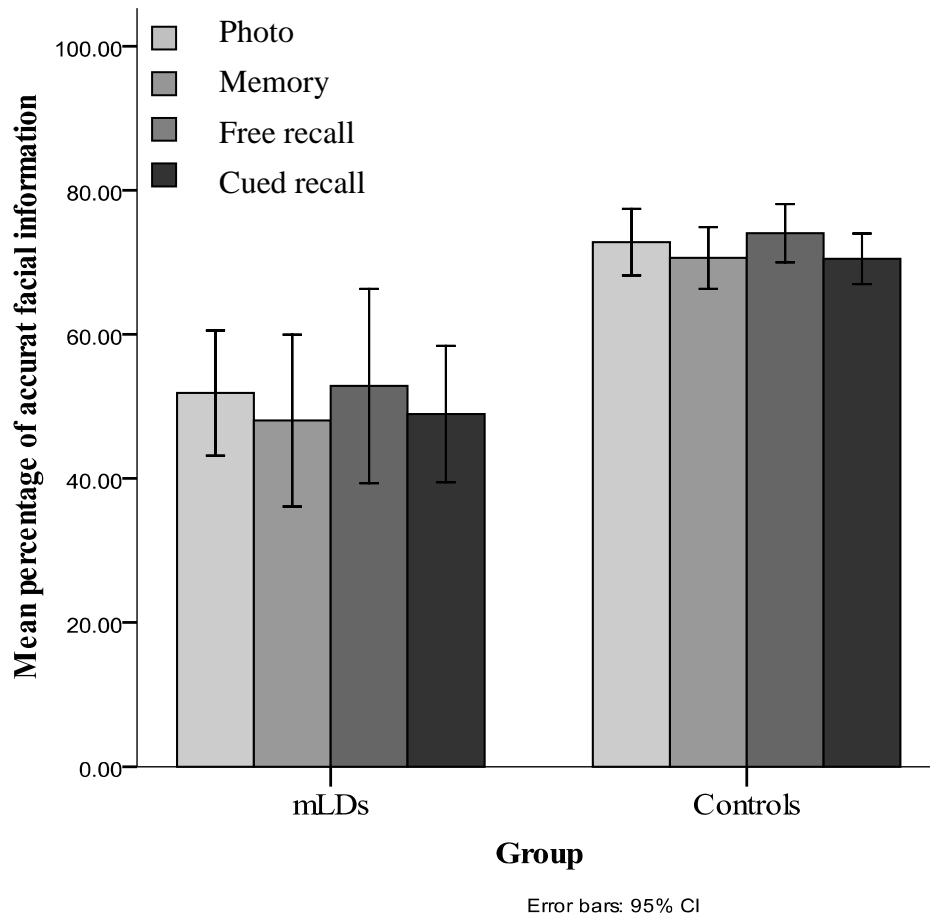


Figure 7.5. Mean percentages of appropriate facial information (+SE) provided by mLD and control participants during the different description modes (photo vs. memory) and recall conditions (free vs. cued). The data are collapsed across conditions.

Correlations between quantity and quality of facial information and mLD

individuals' IQ: Correlations between the verbal-, performance-, and full WASI score (FSIQ-4 score) and the overall quantity and quality of the verbal facial descriptions (collapsed for description mode and recall condition) provided by

participants with mLD were calculated. The analysis revealed that the amount of facial information provided by participants with mLD was significantly correlated with their full, verbal and performance WASI scores, as was the amount of accurate information (quantity of information: full-WASI score: $r = .753$, $p < .001$; verbal-WASI score: $r = .780$, $p < .001$; performance-WASI score: $r = .607$, $p = .001$) (quality of information: full-WASI score: $r = .591$, $p = .001$; verbal-WASI score: $r = .651$, $p < .001$; performance-WASI score: $r = .461$, $p = .014$).

Correlations between quantity and quality of facial information and participants' performance during the main recognition task: Correlations between participants' performance during the main recognition task in Experiment 1 of this Study and the quantity and quality of the verbal facial information provided during the description task in Experiment 2 were calculated. Significant positive correlations were obtained between participants' overall amount of provided facial information ($r = .667$, $p < .001$) and the amount of accurate information mentioned ($r = .665$, $p < .001$). Thus, participants who performed better during the main recognition task in Experiment 1 provided more and also more accurate verbal facial information about the target faces during the description task in Experiment 2 (this was true for both, the control and the mLD group).

Summary

Participants with mLD provided significantly less facial information during the facial description phase than their non-LD counterparts. However, both groups of participants benefited from the same description mode and recall condition. Specifically, all participants mentioned significantly more information when the target photograph was in view compared to when they had to describe it from their memory

alone, and they mentioned significantly more information during the cued recall compared to the free recall. With regard to the quality of facial information, mLD participants mentioned significantly less accurate facial information overall than control individuals. Although not significant, the means indicate that both groups profited from the same description mode and recall condition. Participants provided more accurate information when the target was in view than when they had to describe it from their memory. Moreover, participants provided more accurate information during the free recall than during the cued recall. mLD participants' intellectual functioning, as assessed via the WASI, was significantly correlated with the quantity and quality of facial information provided. Thus, mLD participants with higher intellectual functioning provided more verbal information and more accurate information about the target faces than participants with lower intellectual functioning. Finally, the fact that the same individuals participated during Experiment 1 and 2, allowed us to correlate their face recognition performance, assessed during the main recognition task, with their facial description abilities. It was revealed that the quantity as well as quality of the verbal facial descriptions provided by participants was positively correlated with their face recognition performance. Thus, participants who performed superior on the face recognition task also performed better during the face description task.

7.3.2 Phase 2 Evaluation of facial descriptions

Providing less facial information will not in itself render the description of the faces less forensically useful. It could be the case that all important and relevant information is provided in the mLD descriptions despite them being more brief. To test for this, the descriptions provided by both groups of participants were presented

to an independent sample of participants who were asked to identify the target face after reading the facial description.

Method

Design: A matching task was used to assess the quality of the facial descriptions provided by participants during Phase 1. A 2 x 2 mixed design was used, including one within-subject factor (group: mLD vs. controls) and one between-subject factor (description mode: photo vs. memory).

Participants: Booklets including the matching tasks were handed out to 40 participants, who were all students or staff drawn from the University of Abertay Dundee and the University of Dundee. Thirty-six participants completed the booklets (16 males and 19 females; $M = 34.35$ years; $SD = 10.11$).

Materials: The matching task consisted of booklets comprising the facial descriptions, created during Phase 1 of this experiment. Each facial description was accompanied by a 10 person line-up consisting of the six target faces and four distractor faces. All faces in the line-up were Caucasian males with no distinctive features. All line-ups included the same faces. The images were full-face poses with neutral expressions. The order of the faces in each line-up was randomized. The faces in the line-ups derived from the same database as the images used during Phase 1 of this study. Figure 7.6 depicts an example of the matching task.

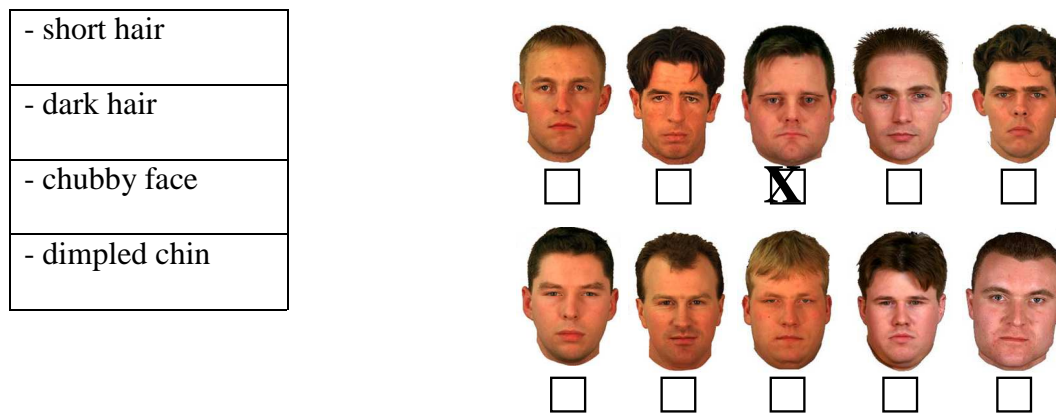


Figure 7.6. *Matching task: In this example, the participant has decided that the third face in the line-up (marked with an X) matches the facial description on the top left of the page most.*

Procedure: Each booklet contained all usable facial descriptions obtained during Phase 1 of the experiment during one of the two description modes. Some of the mLD participants did not provide any relevant facial information at all during the description phase and therefore each booklet contained in total 55 descriptions. Of those 55 descriptions 29 came from the control group and 26 from the mLD group. Each description was displayed on its own page and was accompanied by a 10 person line-up (see Figure 7.6). The order in which the descriptions were presented in the booklets was randomized as was the order in which the distractors and targets were presented in the line-ups. Participants were asked to indicate which of the faces in the line-up best matched the accompanying facial description.

Results

Research Questions: Data analysis addressed the following research questions: First, do participants make more correct matches on the basis of facial descriptions provided by participants with mLD or participants without LD? Second, did the description mode or recall condition have an impact on the overall amount of correct matches participants made? An alpha level of .05 was used for all statistical tests.

Statistical Analysis: The percentages of correct identifications participants made on the basis of descriptions provided by mLD participants and control participants were calculated, and are shown in Table 7.6.

Table 7.6. Mean percentages (+SD) of correct identifications for facial descriptions provided by mLD and control participants during the different description modes (photo vs. memory).

	Percentage correct matches			
	Description mode			
	Memory		Photo	
On the basis of:	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
mLD descriptions	18.80	7.45	16.06	5.48
Control descriptions	35.06	12.64	47.46	11.98

A 2 x 2 mixed ANOVA revealed a significant main effect for group $F(1, 33) = 121.92, p < .001, \eta^2 = .72$. Thus, participants made significantly more correct matches when the facial descriptions were provided by the control group than by the mLD group. The main effect for description mode did not reach significance $F(1, 33) = 530.21, p = .067, \eta^2 = .098$. However, there was a significant interaction effect obtained between group and description mode $F(1, 33) = 12.32, p = .001, \eta^2 = .07$. To elaborate the interaction further, a post hoc test consisting of pair wise comparisons was conducted, using the Bonferroni correction. It was found that the description mode had a significant impact on the amount of correct matches based on control participant's descriptions ($p = .005$), this was however not the case for descriptions provided by mLD participants ($p = .226$). Thus, for descriptions provided by the control group, participants made significantly more correct matches when the descriptions were generated during the photo condition than during the memory condition. For descriptions provided by the mLD group, description mode had no significant impact on the overall amount of correct matches (see Figure 7.7).

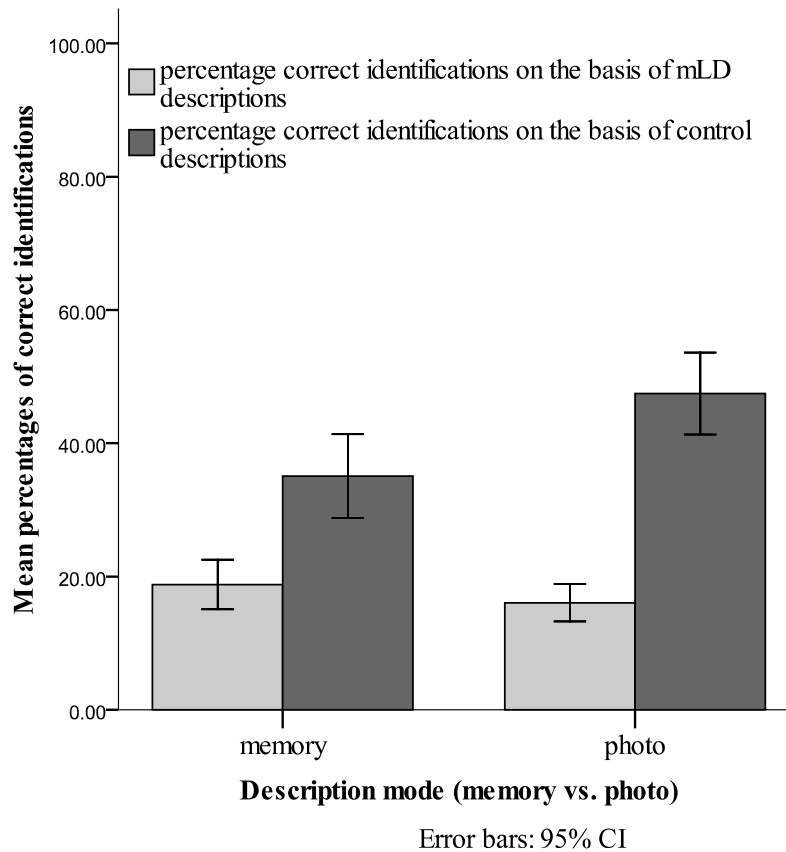


Figure 7.7. Mean percentages of correct identifications (+SE) for facial descriptions provided by mLD and control participants during the different description modes (photo vs. memory).

To examine whether participants made more correct matches on the basis of facial descriptions provided by mLD individuals than would have been expected by chance alone, the critical region of the binomial distribution was calculated. The binomial distribution had a mean of $pn = (1/10)(26) = 2.6$. As stated earlier in this thesis, to determine critical values, both pn and qn must be at least 10. Therefore, it was not possible to calculate the critical region, which would indicate whether an individual was scoring significantly different from chance on this matching task. Instead, the critical region for the whole group of participants was determined to examine whether individuals as a group would make significantly more correct matches on the basis of mLD descriptions than by chance alone. The binomial distribution had a mean of $pn = (1/10)(36 \times 26) = 93.6$ and a standard deviation of

$\sqrt{npq} = \sqrt{936(1/10)(9/10)} = 9.18$. To be significantly different from chance, the group score must be above (or below) the mean by at least $1.96(9.18) = 17.99$. Thus, with a mean of 93.6, the group would need to score above 111.59 ($93.6 + 17.99$) or below 75.61 ($93.6 - 17.99$) to be significantly different from chance performance. The group had a total score of 165 correct matches, which is significantly above chance performance. Thus, although participants had more difficulties accurately identifying the target person out of a 10-person line-up when the facial description derived from a person with mLD, they were nevertheless able, as a group, to make significantly more correct matches than would have been expected by chance alone.

7.4 Discussion

Experiment 1 established that people with mLD performed significantly poorer during the old/new face recognition task than people without LD. However, as a group they were able to manage the task better than expected by chance alone, which indicates that people with mLD are able to remember and subsequently accurately recognise unfamiliar faces. This finding is in line with previous research, which investigated face recognition skills in individuals with LD. Dobson and Rust (1994) demonstrated that people with LD can recognise previously presented stimuli of objects and faces. The ability to remember and later recognise a previously encountered unfamiliar face might be crucial in an eyewitness situation, when a witness or victim has to identify the perpetrator in a police line-up or a mug-shot book. Moreover, this ability forms a prerequisite for the construction of a facial composite, since one has to be able to remember the previously encountered face to be considered capable of creating an accurate composite of it.

The practice tasks revealed that people with mLD needed more training than their non-LD counterparts to understand the instructions and the general procedure of the main recognition task. Since participants with mLD performed at chance level on the first practice task, a second practice task needed to be included to make sure they were able to understand the instructions and the procedure during the main recognition task correctly and to a similar degree as participants without LD. After providing them with sufficient training, participants with mLD were able to complete the main recognition task at better than chance performance. Similar findings were obtained by Dobson and Rust (1994) who found that individuals with LD required more learning trials than individuals without LD to reach the accuracy criterion level of 100% during an old/new recognition task.

To establish whether a higher tendency of acquiescence was responsible for the inferior performance by people with mLD during the main recognition task, the total number of hits and correct rejections were calculated and subsequently analysed. Acquiescence refers to the tendency of an individual to respond to questions in an affirmative way, regardless of the actual content of the questions (Gudjonsson, 1990). Previous studies found significant negative correlations between intelligence and acquiescence (Clare & Gudjonsson, 1993; Gudjonsson, 1990). During the current recognition task, participants with mLD did not show an increased tendency of acquiescence. They correctly rejected the new faces at a similar degree as they made hits for the old faces. This finding implies that people with mLD can withstand acquiescence when questioned in a non-suggestive manner.

A possible limitation of the main face recognition task in Experiment 1 may be that the same stimuli were used during the learning and test phase. It could be argued that the task rather assessed participants' picture recognition performance than

their face recognition skills. However, given that participants with mLD took part in this Study, it was deemed important to make the task not too difficult and therefore the same stimuli were used during the learning and test phase. Furthermore, the background of all stimuli was grey and the faces did not possess any distinctive features, which may have prevented participants from egenging in picture recognition and promoted face recognition instead.

During Experiment 2, verbal facial descriptions of people with mLD were evaluated and compared with those provided by individuals without LD. In the evaluation phase, it was established that an independent sample of participants, not involved in the face description phase, had significantly more difficulties identifying the target face out of a 10 person line-up when the description originated from a person with mLD than when it derived from an individual without LD. Several factors might have been responsible for the poorer quality of the verbal facial descriptions. First, the descriptions from mLD participants were less detailed than those provided by the non-LD participants. In fact, it should be noted that a minority of participants with mLD (four individuals) did not provide any relevant facial information about the target faces at all. This finding is in agreement with earlier research, which revealed that individuals with LD provided less information about an observed event compared to individuals without LD (Agnew & Powell, 2004; Brown & Geiselman, 1990; Henry & Gudjonsson, 2004; Michel et al. 2000; Milne, 1999; Perlman et al., 1994). Second, participants with mLD not only mentioned fewer items of facial information but also less accurate information. The fact that this was found in the photo present condition, suggests that the vocabulary of mLD participants for describing a face is limited and not as elaborate as the vocabulary of people without LD. Although, facial descriptions from participants with mLD were overall poorer, in quantity as well as in

quality, participants were still able to perform the matching task at a better rate than would have been expected by chance alone. This demonstrates that at least some of the facial descriptions provided by people with mLD were accurate enough to correctly identify the target face out of 10-person line-up on the basis of them.

Intriguingly, although people with mLD performed at a lower level overall, they benefited from the same introduced measures as people without LD. With regard to question format, both participant groups mentioned significantly more facial information during the cued recall compared to the free recall. This result is consistent with previous research that investigated the impact of question format on eyewitness accounts (see Memon & Bull, 1999 for a concise overview). The employed question format had no significant impact on the accuracy of the provided facial information. However, trends indicated that both participant groups provided more accurate facial information during the free recall than during the cued recall. This again is in line with previous research, which has revealed that eyewitness accounts of both LD and non-LD individuals become more inaccurate when a more specific questioning approach is applied (Agnew & Powell, 2004; Dent, 1986; Henry & Gudjonsson, 1999; Michel et al., 2000).

With regard to the intellectual functioning of people with mLD, which was assessed via the WASI, a significant relationship was revealed between the full WASI score and the performance of participants with mLD on the main recognition task. This indicates that mLD individuals with higher intellectual skills can better deal with increased task demands, such as a higher amount of presented stimuli, during a forced choice recognition task. Furthermore, significant correlations were obtained during the second experiment, between mLD individual's WASI scores and the quantity and quality of the provided facial descriptions. This finding suggests that IQ scores,

obtained with psychometric tests, could be used as predictive measure for mLD individuals' ability to completely and accurately describe an unfamiliar face.

Furthermore, participants with superior face recognition abilities also performed better during the face description task, as evidenced by significant positive correlations. This finding appears reasonable since according to the Bruce and Young model (1986), recognising unfamiliar faces involves comparing the present encoded facial representation with the pictorial code which was formed during the initial encounter. It follows that describing an unfamiliar face, particularly from memory, requires the individual to recall the facial information from the stored pictorial code. Peoples' ability to encode faces in a very elaborate way, resulting in a detailed and superior pictorial code, may therefore not only be beneficial during tasks involving face recognition but also during tasks involving verbally describing faces. In view of this, it could be argued that an easy and quick-to-deliver face recognition task, as the one applied during the current study, may be useful in a legal setting to determine basic eyewitnesses' abilities, including the accurate recognition and/or description of the perpetrator's face.

Taken together, these results suggest that there is initial evidence that people with mLD are consistently poorer in performance on face recognition and recall tasks, fitting with the generally held layman's view that they might be less reliable eyewitnesses (Peled et al., 2004; Stobbs & Kebbell, 2003). However, there is also evidence that people with mLD can perform those tasks better than would have been expected by chance alone and exhibit variability in performance dependent on the demands of the task. This suggests they might benefit from measures introduced to facilitate performance, such as prior training trials and different forms of question format. After having established that individuals with mLD are able to accurately

recognise and describe previously seen unfamiliar faces, it seems reasonable to examine their face recognition and description abilities in a more applied setting, such as during the construction of a facial composite image. The next chapter describes an experimental study investigating the ability of people with mLD to use E-FIT, a facial composite system frequently used by the UK police.

Chapter 8

Study 3: The Efficiency of E-FIT with mLD witnesses

This chapter describes Study 3, which investigates the efficiency of the E-FIT system with mLD witnesses. Study 2 established that individuals with mLD have limited verbal abilities when describing faces. The present study investigates whether this might act as a barrier to them creating accurate facial composites with current facial composite systems. Two participant groups, one with mLD and one without LD, are required to use the E-FIT system to construct facial composites of unfamiliar faces. The resulting composites are subsequently evaluated by an independent sample of participants. The obtained results provide an insight into difficulties that people with mLD might experience when using the E-FIT system.

8.1 Introduction

As described in detail in Chapter 2, people with LD are more likely to be witnesses to crime or victims of crimes than other members of the general population (Kebbel & Hatton, 1999). This suggests that they are more likely to be placed in the situation of having to provide a description of a perpetrator's face to the police. However, as emphasised in Chapter 3, people with LD are also more likely to be excluded from general criminal justice procedures, such as creating a facial composite image, since research has repeatedly shown that they are regarded by the majority of people as less credible and accurate witnesses (Peled, Iarocci & Connolly, 2004; Stobbs & Kebbell, 2003). Despite the common scepticism regarding the ability of witnesses with LD to provide reliable evidence, it is surprising that to date no study has investigated their ability to construct facial composites of unfamiliar faces.

During the previous experimental investigation (Study 2), it was revealed that individuals with mLD differed from their non-LD peers in their face recognition and description skills. Participants with mLD displayed difficulties in recognising previously seen unfamiliar faces and their verbal facial descriptions contained only sparse information. Moreover, participants with mLD used different terminology to describe the target faces than individuals without LD. The terminology used by the non-LD group was much more in agreement with the terms provided by the Aberdeen Index in the E-FIT system. As explained in Chapter 4, composites created with E-FIT are based on verbal descriptions of witnesses gathered through the CI at the beginning of the E-FIT construction process. The provided facial information is then entered by the operator into the Aberdeen Index, which drives an algorithm that automatically selects the most fitting features from the data base. Language deficiencies, as observed during Study 2, might therefore have a detrimental effect on the whole construction process and the quality of the subsequent composite image.

The aim of this investigation was to examine the ability of people with mLD to use facial composite systems currently utilised by the UK police. In particular, their ability to work with the E-FIT system was explored. Two groups of participants, one with mLD and one without LD were required to use E-FIT with and without the target faces in view. These two description modes (photo vs. memory) were included in order to investigate further whether any arising difficulties during the composite construction process in individuals with mLD are due to memory or language deficits, or a combination of both. The study consisted of two phases, the description and the evaluation phase. During the description phase, E-FIT composites were constructed on the basis of the facial descriptions provided by the participants. During the

subsequent evaluation phase, the resultant composites were evaluated by an independent sample of participants using a matching task and a likeness rating task.

8.2 Phase 1: Composite construction

8.2.1 Method

Design

A 2 (group: mLD vs. controls) x 2 (description mode: photo vs. memory) between-subjects design was employed.

Participants

Overall, 60 participants took part in this study (25 males and 35 females). Of those, 30 were people with mLD (19 - 68 years; $M = 43.17$ yrs; $SD = 12.22$; WASI: FSIQ-4 score: $M = 57.97$, $SD = 3.63$; WASI: verbal score: $M = 57.13$, $SD = 2.76$; WASI: performance score: $M = 64.47$, $SD = 5.08$) and 30 were people without LD (19 - 48 years; $M = 28.67$ yrs; $SD = 7.58$). Participants with mLD were recruited from social day care centres in and around the Dundee area. Control participants were recruited from the student and staff body of the University of Abertay Dundee. General intellectual functioning of participants with mLD was assessed using the WASI. No participants recruited to be part of the mLD group were excluded from the study on the grounds of their IQ scores being outside of the desired range.

Materials

Facial stimuli: Five static full-face photographs of unfamiliar Caucasian males were used as stimuli. A meta-analysis conducted by Shapiro and Penrod (1986) evaluated 13 face perception studies and found only minor changes in participants' behaviour

between studies involving live or video-taped stimuli and those presenting static photographs. This indicates that a static stimulus is unlikely to affect the composite quality in any detrimental way. The targets were selected from a larger sample of unfamiliar face photographs. All target faces were presented without spectacles or other distinguishing marks, the facial expressions were neutral and all faces were shown from the front. Each photo was 600 x 800 pixels in size. The facial stimuli derived from a database created by the UK Home Office PITO and are depicted in Appendix 3.

Facial composite system: Aspley E-FIT version 6.0 for Windows was used to create facial composites. The E-FIT software was run on a Toshiba Satellite Pro A200 laptop running Windows XP. The experimenter completed a training course in E-FIT.Net 6.02 organized by Vision Metric.

Paint Program: Micrografx Picture Publisher 8 for Windows was used to modify and enhance, as necessary, the quality of the resulting composites in a number of ways (e.g. adding stubble and/or age lines).

Procedure

A pilot study was conducted with three participants (one control participant and two mLD participants) to enable the experimenter to practice and to test the oral instructions, the stimuli, the software programs and the feasibility of the procedure as a whole. The pilot sessions were video taped and prior to the actual experiment presented to the supervisory team of the experimenter for inspection to make sure that the procedure was standardised.

During the appropriate and actual experiment, the procedure for both participant groups (mLDs and controls) was the same. Composites were constructed

through two different description modes: photo versus memory. In the photo condition, the target face remained in view whilst the experimenter worked with the participant to create the composite image. In the memory condition, the composite was created from the memory of the participant alone. All experimental sessions were video-taped. To increase the ecological validity of the study, during the memory condition, targets were presented to the participants in the morning (morning session), and the composites were constructed in the afternoon (afternoon session). This created a delay of at least 3 hours between the actual presentation of the target and the construction of the composite (a similar delay was used in a study conducted by Frowd et al., 2005a).

At the beginning of the experimental session, the experimenter provided each participant with a short outline of the procedure and an oral explanation combined with a demonstration of how the E-FIT software package operates. Participants were asked to randomly select one of the five target faces, by pulling the picture of the target face out of an envelop. The identity of the target face was kept a secret from the experimenter at all times during the construction of the composite to avoid any possible biases. Participants in the memory condition viewed the target face for one minute and were asked to come back later after a three hour delay. Participants in the photo condition were allowed to look at the target face as often and for as long as they wished, during the composite construction.

The procedure used to construct the composites followed the recommendations of the ACPO(S) Working Group for Facial Identification Guidelines (2009) to police operators. First, the CI was administered and a full verbal description of the target face was obtained from the participant. Second, the obtained facial information was entered into the Aberdeen Index of the E-FIT system by the

experimenter. The resulting composite was then presented to the participant who was allowed to make changes to features by changing their size or position, or by scrolling through alternative features. When a sufficient likeness emerged, further fine-grained changes were made if required by the participant, with the use of the Microsoft Paint program. The experimenter finished the composite construction process when the participant stated that he/she was satisfied with the likeness of the resulting composite (see Figure 8.1 for a screenshot of the E-FIT system).

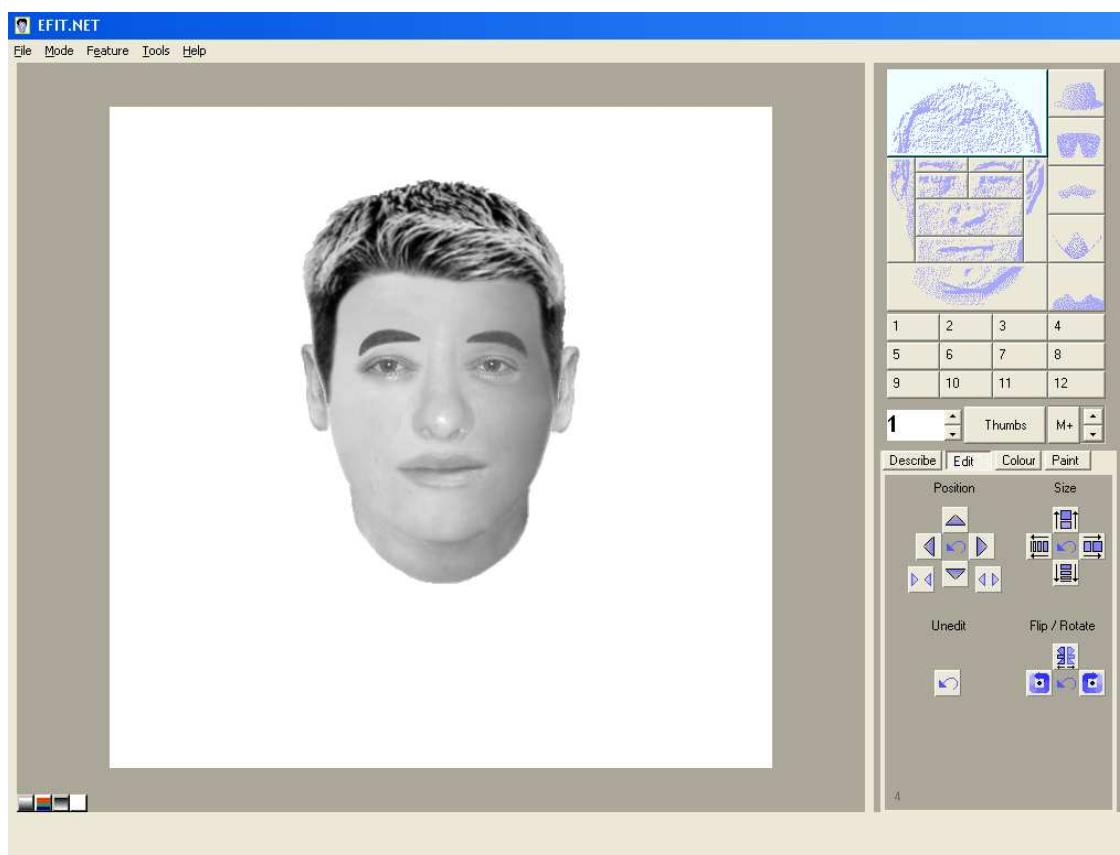


Figure 8.1. Screenshot of the E-FIT system.

Scoring

Quantity of facial information during the CIs: To examine the quantity of facial information participants provided during the CIs, all gathered verbal descriptions were transcribed and the overall amount of facial information mentioned was counted. The following scoring procedure was used: the response *black hair* was coded as one

piece of facial information, *the response long, black hair* was coded as two pieces of facial information and the response *long, curly black hair* was coded as three pieces of information. Each piece of facial information provided by the participant was counted only once during each recall condition. Each recall condition was coded independently.

Non-verbal behaviour during the composite construction: To investigate whether mLD participants use more gestures to facilitate communication during the composite construction process than members of the control group, the overall number of times each participant pointed to his/her own face and the number of times a participant pointed to the monitor, to aid the description of the target face, was counted.

Number of features changed: The total amount of facial features (out of 7: hair, eyebrows, eyes, nose, mouth, face shape and ears) requested to be changed during the composite construction phase was counted for each participant. Changes included: Participant wants to see other exemplars of the feature, participant wants to change the size or the position of the feature, and participant wants to make changes in the Paint Program.

Number of exemplars: The total number of exemplars presented to each participant before he/she expressed for the first time being satisfied with one was counted.

Number of acceptances and rejections: The number of times a participant accepted an exemplar and the number of times a participant rejected an exemplar of a specific feature was counted and subsequently overall percentages of acceptances and rejections were calculated for each participant.

Coder reliability

Intra-coder reliability: To establish the intra-coder reliability, the person who coded the CI transcripts and the composite construction sessions for the first time coded 10 randomly selected CI transcripts and composite construction sessions a second time. This was done to see how consistent the coder's decisions were. Correlational analysis revealed significant levels of agreement between the two coding occasions for all variables (all $ps < .001$).

Inter-coder reliability: To determine the inter-coder reliability, two coders coded the transcripts of the CIs and the video-taped composite construction sessions independently. The correlational analysis showed a significant level of agreement for all variables between the two independent coders on the basis of a random sample of 10 sessions (all $ps < .001$).

8.2.2 Results

Research Questions

The analysis of data examined the following research questions: First, do individuals with mLD differ with regard to the amount of verbal information they provide about the target face during the CI from participants without LD? Based on previous research findings on eyewitness accounts of individuals with LD about a to-be-observed event (Agnew & Powell, 2004; Milne, Clare & Bull, 1999), it was hypothesised that people with mLD would provide less facial information during the CIs than members of the control group. Second, do participants with mLD differ with regard to their non-verbal behaviour they show during the composite construction process from participants without LD? No directional hypothesis were made here,

however, it was assumed that individuals with mLD would display more non-verbal behaviours than people without LD, to compensate for their communication deficiencies. Third, how much time do participants require to create a composite image? Fourth, are there differences between the two participant groups in the amount and type of facial features they want to change during the composite construction process? And finally, do individuals with mLD show a higher tendency of acquiescence compared to non-LD participants during the composite construction? Earlier research has demonstrated that people with LD show a high tendency of acquiescence (Gudjonsson & Henry, 2003; Henry & Gudjonsson, 1999; Henry & Gudjonsson, 2003; Milne, Clare & Bull, 2002), therefore it was hypothesised that participants with mLD will show behaviours which are in agreement with this research finding, such as being more easily satisfied with the composite image and requesting fewer changes to be made to enhance the likeness of the composite. An alpha level of .05 was used for all statistical tests.

Verbal information obtained during the CI

The video-taped CIs were subsequently transcribed to enable an investigation of the total number of facial information provided by the participants. See Table 8.1 for means.

Table 8.1. *Mean number of facial information (+SD) provided by mLD and control participants in the CIs during the different description modes.*

Group	Description mode			
	Photo		Memory	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
mLDs	42.53	32.67	29.13	15.56
Controls	105.80	46.11	88.13	24.75

A 2 (group: mLD vs. controls) x 2 (description mode: photo vs. memory) between-subjects design ANOVA was carried out. The Levene's test was significant ($p = .01$), indicating that the variances were significantly different across groups. One of the assumptions of the ANOVA had been therefore violated. As a result, in line with Hinton, Brownlow, McMurray and Cozens's (2004) recommendations, the data were log transformed (\ln) and then re-analysed. The transformation helped to stabilise the data and the Levene's test was no longer significant ($p = .14$). The ANOVA revealed only one significant main effect for group, $F(1, 56) = 64.58, p < .001, \eta^2 = .48$. Participants with mLD reported significantly less facial information during the CI than control participants (see Table 8.1). The main effect for description mode $F(1, 56) = 2.24, p = .140, \eta^2 = .03$ and the interaction between group and description mode $F(1, 56) = .145, p = .482, \eta^2 < .00$ did not reach significance. Nevertheless, the means pointed into the expected direction, both participant groups mentioned more information during the photo than the memory condition (see Table 8.1).

Non-verbal information present during the composite construction

Pointing to their own face: To examine whether there was a difference between the two groups of participants (mLDs vs. controls) in the total amount of times they pointed to their own face to facilitate the description of the target face during the composite construction process, a 2 (group: mLDs vs. controls) x 2 (description mode: photo vs. memory) between-subjects design ANOVA was conducted. The dependent variable was the number of times the participants pointed to their face to facilitate the composite construction. The Levene's test was significant ($p = .007$). Consequently, the data were log transformed (\ln) and re-analysed. The transformation stabilised the data and the Levene's test was no longer significant ($p = .992$). The

ANOVA revealed only one significant main effect for group $F(1, 56) = 11.98, p = .001, \eta^2 = .17$. Thus, participants with mLD pointed significantly less often at their own faces to facilitate the description of the target face during the composite construction process than members of the control group (Table 8.2 shows the means). The main effect for description mode $F(1, 56) = .43, p = .51, \eta^2 = .00$ and the interaction between group and description mode $F(1, 56) = 1.71, p = .20, \eta^2 = .02$ were non-significant.

Table 8.2. *Mean number of times mLD and control participants pointed to their own faces (+SD) during the two description modes.*

Group	Description mode			
	Photo		Memory	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
mLDs	7.93	6.98	4.80	4.60
Controls	15.80	15.72	17.33	16.58

Note. Mean values and standard deviations of the not log transformed data

Pointing to the monitor: To investigate whether there was a difference between the two groups of participants (mLDs vs. controls) in the total amount of times they pointed to the monitor to facilitate the description of the target face a 2 (group: mLDs vs. controls) x 2 (description mode: photo vs. memory) between-subjects design ANOVA was conducted. The dependent variable was the number of times the participants pointed to the monitor during the composite construction phase. The Levene's test was again significant ($p < .001$). In line with previous practice the data were log transformed (ln) and re-analysed, however the Levene's tests remained significant ($p < .001$). Since the parametric assumptions were not met, a non-parametric test was used to analyse the data. There is no non-parametric equivalent for a two factor ANOVA, therefore, the data were split and analysed separately for the photo and memory condition with a Mann-Whitney test. The test revealed that during

both description modes, control participants pointed significantly more frequently to the monitor than mLD participants to assist the composite construction (mLDs: $Mdn = 8.5$, controls: $Mdn = 53.5$) (photo: $U = 18$; $p < .001$, $r = -.71$; memory: $U = 23$; $p < .001$, $r = -.68$). To examine further whether description mode had an effect on the number of times participants pointed to the monitor, the description mode data was split and analysed separately for the mLD and control group. For both participant groups, the Mann-Whitney test revealed no significant difference in the number of times they pointed to the monitor between the photo and the memory condition (mLDs: photo: $Mdn = 10$, memory: $Mdn = 8$; controls: photo: $Mdn = 64$, memory: $Mdn = 47$) (mLDs: $U = 228.50$; $p = .87$, $r = -.03$; controls: $U = 71.00$; $p = .08$, $r = -.31$) (see Table 8.3 for the means and standard deviations). This finding is consistent with the one obtained for pointing to the face.

Table 8.3. Mean number of times mLD and control participants pointed to the monitor (+SD) during the two description mode conditions.

Group	Description mode			
	Photo		Memory	
	\underline{M}	\underline{SD}	\underline{M}	\underline{SD}
mLDs	14.33	16.41	9.60	8.76
Controls	60.27	30.16	41.80	24.41

Duration of the composite construction

To examine whether there were any differences in the amount of time participants spent engaging with the E-FIT software, the duration (in minutes) of the composite construction was calculated for each participant. A between-subjects design ANOVA was carried out. The Levene's test was significant ($p = .01$), therefore, the data were log transformed (ln) and reanalysed. After the data transformation the Levene's test was no longer significant ($p = .96$). The ANOVA revealed a significant main effect

for group $F(1, 56) = 53.14, p < .001, \eta^2 = .47$. Thus participants with mLD spent significantly less time ($M = 19.23$ minutes, $SD = 11.15$) with the construction of the composites than participants from the control group ($M = 49.17$ minutes, $SD = 21.40$). The main effect for description mode $F(1, 56) = 1.61, p = .21, \eta^2 = .01$ and the interaction effect between group and description mode $F(1, 56) = .00, p = .96, \eta^2 = .00$ were not significant.

Duration spent in E-FIT program vs. Paint program: To examine whether the obtained difference in the duration of the composite constructions between the two participant groups was due to the fact that participants without LD required more time in general with the composite creation or whether they were spending less time amending the composite image further using the Paint software than participants with mLD, a 2 (group: mLDs vs. controls) x 2 (description mode: photo vs. memory) x 2 (program: E-FIT only vs. Paint) mixed ANOVA was carried out. The between-subject factors were group and description mode, while program (E-FIT vs. Paint) was the within-subjects factor. The Levene's test for the time spent in the Paint package was significant (E-FIT: $p = .08$; Paint: $p = .01$). The data were log transformed (ln) and then re-analysed. The transformation stabilised the data and the Levene's test was no longer significant (E-FIT: $p = .729$; Paint: $p = .076$). Two significant main effects were obtained. A main effect for program was revealed $F(1, 56) = 19.26, p < .001, \eta^2 = .21$, indicating that on average individuals spent more time in E-FIT than in Paint. Another main effect for group was observed $F(1, 56) = 63.68, p < .001, \eta^2 = .52$, demonstrating that control participants spent significantly more time in the E-FIT program and in the Paint program than mLD participants (see Table 8.4 for relevant means and standard deviations). The main effect for description mode did not reach significance $F(1, 56) = 1.64, p = .21, \eta^2 = .01$.

Table 8.4. Mean durations of time (+SD) spent in the E-FIT program and the Paint program for both participant groups.

Group	Duration			
	E-FIT		Paint	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
mLDs	12.78	7.67	6.45	6.45
Controls	25.55	13.19	23.62	12.43

A significant interaction effect was observed between program and group $F(1, 56) = 12.53, p = .001, \eta^2 = .14$. To explore this interaction effect further, a post hoc test consisting of pair wise comparisons was conducted, using the Bonferroni correction. The post hoc test revealed a significant difference in the duration of time spend in E-FIT and Paint for the mLD group ($p < .001$), however, no such significant effect was present for the control group ($p = .38$). Thus, participants with mLD spent considerably less time in the Paint program than with the E-FIT program, whereas non-LD participants spent equal amounts of time in the E-FIT and the paint program. Figure 8.2 shows the interaction effect. The other two two-way interactions (program*description mode $F(1, 56) = 1.96, p = .167$ and group*description mode $F(1, 56) = .05, p = .825, \eta^2 = .00$) and the three-way interaction (program*group*description mode $F(1, 56) = .10, p = .747, \eta^2 = .00$) were not significant.

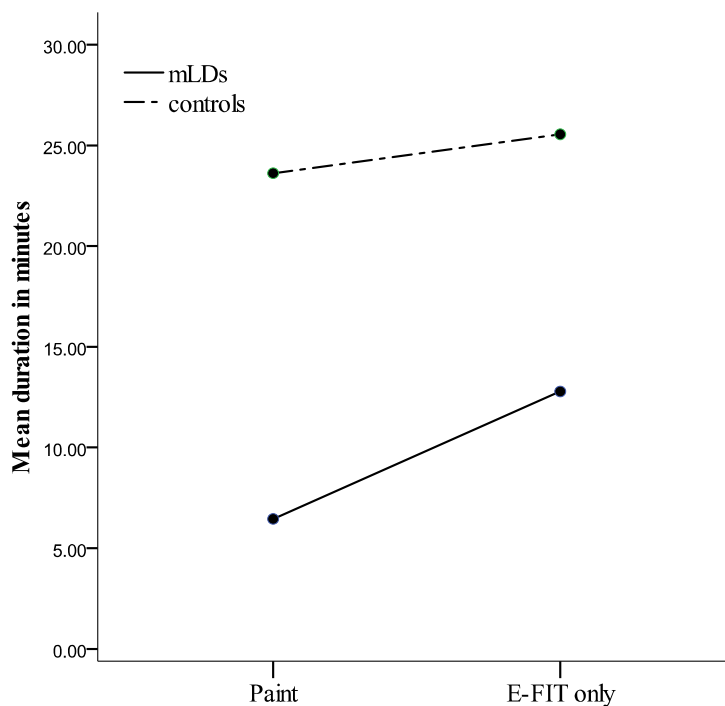


Figure 8.2. Mean durations of time (in minutes) that both participant groups spend in the Paint program and the E-FIT program.

Features requested to be changed during the composite constructions

Amount of facial features changed: To investigate whether the two participant groups differed in the total amount of facial features they wanted to have changed during the composite construction, the total amount of features (out of 7: hair, eyebrows, eyes, nose, mouth, face shape and ears) requested to be changed was calculated for each participant. A 2 (group: mLDs vs. controls) x 2 (description mode: photo vs. memory) between-subjects design ANOVA was conducted to analyse the data. The Levene's test was significant ($p < .001$). Therefore, the data were log transformed (\ln) and re-analysed. The data transformation had no effect and the Levene's test remained significant ($p < .001$). Consequently, the data were split and analysed separately for the photo and memory condition with the Mann-Whitney test to investigate the effect of group. The test revealed that during both description modes, mLD participants

requested to change significantly fewer features than members of the control group (mLDs: $Mdn = 5$, controls: $Mdn = 7$) (photo: $U = 18$, $p < .001$, $r = -.71$; memory: $U = 23$, $p < .001$, $r = -.68$) (see Table 8.5). To investigate further whether description mode had an effect on the amount of features requested to be changed by the participants, the data were split and analysed separately for the mLD and control group, with description mode as the independent variable. For both participant groups, the Mann-Whitney test found no significant difference in the total amount of features requested to be changed between the photo and the memory condition (mLDs: photo: $Mdn = 5$, memory: $Mdn = 6$; controls: memory: $Mdn = 7$, photo: $Mdn = 7$) (mLDs: $U = 108.50$, $p = .87$, $r = -.03$.; controls: $U = 71$, $p = .089$, $r = -.31$) (Table 8.5).

Table 8.5. Mean number of facial features (out of seven) requested to be changed (+SD) for both participant groups during the two different description modes.

Group	Description mode			
	Photo		Memory	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
mLDs	4.67	1.68	5.20	1.97
Controls	6.67	0.62	6.73	0.46

Type of facial features changed (external vs. internal): To examine whether people with mLD wanted to change different types of facial features than participants without LD, the total amount of external (hair, ears, face shape) and internal (eyebrows, eyes, nose, mouth) features requested to be changed was calculated for each participant. A 2 (group) x 2 (description mode) x 2 (feature) mixed design ANOVA was carried out. The two between-subject factors were group (mLDs vs. controls) and description mode (photo vs. memory) and the within-subject factor was feature (external vs. internal). The Levene's test was significant for the amount of internal features mentioned ($p > .001$). Consequently, the data were log transformed

(ln) and subsequently reanalysed. However, the Levene's tests remained significant ($p < .001$).

Again, the data were collapsed across description mode, and a Mann-Whitney test was carried out. The independent variable was group and the dependent variables were the total amount of changed external features and internal features. The analysis revealed that mLD participants wanted to change significantly less external features ($M = 2.27$, $SD = .64$, $Mdn = 2$) as well as internal features ($M = 2.67$, $SD = 1.45$, $Mdn = 3$) than members of the control group ($M = 2.70$, $SD = .54$, $Mdn = 3$; $M = 3.87$, $SD = .43$, $Mdn = 4$ for external and internal features retrospectively); (external features: $U = 282.50$; $p = .005$, $r = -.48$; internal features: $U = 214$; $p < .001$, $r = -.67$). To investigate whether there was a difference between the two participants groups in relation to what type of facial feature they wanted to have changed a Wilcoxon signed-rank test was conducted for each participant group separately. For the mLD group, there was no significant difference between the number of internal features and external features participants wanted to have changed ($T = 100.50$, $p = .14$, $r = -.27$). For the control group, the difference was significant ($T = 0$, $p = .017$, $r = -.89$). Specifically, control participants wanted to change significantly more internal than external facial features (Table 8.6). To ensure that any significant differences were not due to the fact that there are different amounts of external and internal features, namely three external and four internal ones, percentages were calculated and the statistical analysis was repeated. Similar results were obtained as before with the raw data, i.e. a significant difference between the number of internal and external facial features requested to be changed by the control group (internal: $Mdn = 100\%$, external: $Mdn = 100\%$) ($T = 95.50$, $p = .02$, $r = -.43$). No significant difference for the

type of feature requested to be changed was obtained for the mLD group (internal: $Mdn = 75\%$, external: $Mdn = 66.6\%$) ($T = 5$, $p = .19$, $r = -.43$).

Finally, to examine whether the description mode had a significant effect on the type of feature participants wanted to change more frequently, a Mann-Whitney test was conducted for the mLD group and the control group separately. mLD participants wanted to change external ($U = 101$, $p = .59$, $r = -.09$) and internal facial features ($U = 87$, $p = .27$, $r = -.20$) to an equal amount during both description modes (external: photo: $Mdn = 2$, memory: $Mdn = 2$; internal: photo: $Mdn = 3$, memory: $Mdn = 4$). The same holds for the control group (external: photo: $Mdn = 3$, memory: $Mdn = 3$; internal: photo: $Mdn = 4$, memory: $Mdn = 4$) (external features: $U = 96$, $p = .374$, $r = -.16$; internal features: $U = 104.50$, $p = .52$, $r = -.12$). Table 8.6 shows the means and standard deviations.

Table 8.6. Mean number of external and internal facial features requested to be changed (+SD) by the two participant groups during the different description modes.

Group	Description mode							
	Photo				Memory			
	External Features ^a		Internal Features ^b		External Features ^a		Internal Features ^b	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
mLDs	2.20	.68	2.47	1.36	2.33	.62	2.87	1.55
Controls	2.80	.41	3.93	.26	2.60	.63	3.80	.56

^a out of three; ^b out of four

Number of exemplars presented

For each participant the average number of exemplars seen before choosing one was calculated. A 2 (group) x 2 (description mode) between-subjects design ANOVA was carried out to investigate whether mLD participants wanted to see fewer exemplars before choosing one than control participants. The Levene's test was significant ($p =$

.01). The data were log transformed (ln) and reanalysed. After the log transformation, the Levene's test was no longer significant ($p = .32$). Two significant main effects were observed. There was a significant main effect for group $F(1, 56) = 10.58, p = .002, \eta^2 = .14$. Participants with mLD asked to see significantly fewer exemplars before being satisfied than members of the control group. Another significant main effect for description mode was found $F(1, 56) = 5.65, p .021, \eta^2 = .07$. Participants wanted to see significantly less exemplars during the memory condition than during the photo condition before choosing one (see Table 8.7 for means and standard deviations). The interaction effect between group and description mode was not significant $F(1, 56) = 3.35, p = .072, \eta^2 = .04$.

Table 8.7. *Mean number of exemplars participants requested to see before choosing one (+SD) during the two description modes.*

Group	Description mode			
	Photo		Memory	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
mLDs	4.16	2.79	2.04	1.24
Controls	4.71	2.33	4.27	1.85

Tendency to acquiesce

To investigate whether participants with mLD tend to show a tendency to acquiesce when viewing the presented exemplars, the overall percentage of accepted exemplars during the composite construction phase was calculated for each participant. A 2 (group: mLDs vs. controls) x 2 (description mode: photo vs. memory) between-subjects design ANOVA was carried out. The dependent variable was the percentage of acceptances. The Levene's test was significant ($p < .001$), therefore the data were log transformed (ln) and re-analysed. The Levene's test remained significant ($p < .04$). Consequently, the data were split first for description mode and two Mann-

Whitney tests were carried out for each description mode condition. The independent variable was group and the dependent variable was percentage acceptances. For the photo condition, a significant difference was obtained between the mLD and the control group in the overall percentage of acceptances (mLDs: $Mdn = 36\%$, controls: $Mdn = 22\%$) ($U = 55.50$; $p = .02$, $r = -.43$). Thus, mLD participants accepted on average significantly more presented exemplars than members of the control group. On average, the mLD group accepted 38% of the present exemplars whereas the control group only accepted 20.40%. For the memory condition, the same result was revealed (mLDs: $Mdn = 69\%$, controls: $Mdn = 26\%$) ($U = 28.50$; $p < .001$, $r = -.64$). Again, mLDs accepted on average significantly more presented exemplars than members of the control group, specifically, 63.6% and 27.2% for mLD and control participants retrospectively.

To investigate further whether there are differences between the two description modes and the overall percentage of accepted exemplars, the data were split for group and two Mann-Whitney tests were conducted for each participant group (mLDs & controls). The independent variable was this time description mode and the dependent variable was again percentage acceptances. For the mLD group, there was a significant difference between the photo and the memory condition in the overall percentage of accepted exemplars (photo: $Mdn = 36\%$, memory: $Mdn = 69\%$) ($U = 61$; $p = .03$, $r = -.39$). Thus, mLD participants accepted on average more exemplars during the memory condition than during the photo condition. For the control group no significant difference for the photo and the memory condition in the overall amount of accepted exemplars was obtained (photo: $Mdn = 22\%$, memory: $Mdn = 26\%$) ($U = 72$; $p = .09$, $r = -.31$). Figure 8.3 shows the mean percentages.

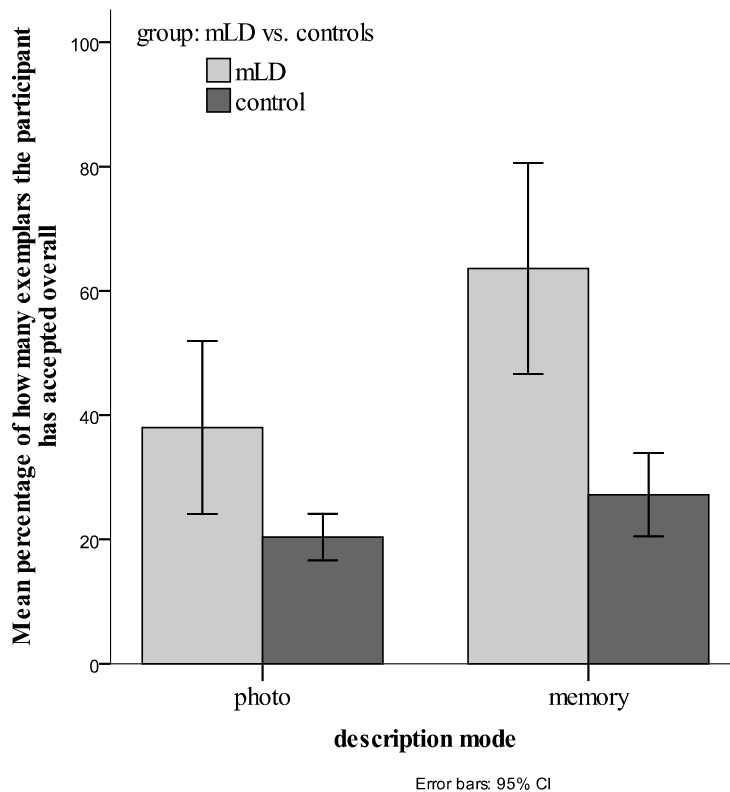


Figure 8.3. Mean number of percentages of accepted exemplars (+SE) by the two groups of participants during the two different description modes.

Relationship between WASI scores and performance

Correlations between the WASI scores of participants with mLD and their performance during the composite construction phase were calculated. Significant positive correlations were obtained between the mLD individuals' full WASI score and the durations of time they spend in the E-FIT program ($r = .53, p = .003$) and the Paint program ($r = .38, p = .03$) to create the composite image. Furthermore, the full WASI score was also significantly positively correlated with the amount of facial information obtained during the CI ($r = .53, p = .003$). When dividing the full WASI score into the verbal and performance WASI scores, the correlational analysis revealed a significant positive association between the verbal WASI score and the duration of time mLD participants spend in the E-FIT program ($r = .37, p = .04$). Furthermore, the performance WASI score was significantly positively correlated

with the duration of time mLD individuals spent in the E-FIT program ($r = .45, p = .01$) and the Paint program ($r = .48, p = .007$) and the amount of facial information provided during the CI ($r = .62, p < .001$). No other significant correlations were obtained between mLD individuals' WASI scores and their performance during the E-FIT construction.

Summary of main findings

During the E-FIT construction phase participants with mLD differed significantly from individuals without LD in several ways. First, during the CIs, participants with mLD reported significantly less verbal facial information than control participants. A similar pattern was obtained with regards to non-verbal facial information provided by the participants in the form of pointing to their own face or to the monitor to facilitate the construction of the composite image and to assist the experimenter. Participants with mLD pointed significantly less often to their own face and to the monitor than members of the control group. Thus, overall the mLD group provided the experimenter with both less verbal as well as non-verbal information about the target face during the composite construction. Second, participants with mLD spent significantly less time creating the composite images than individuals from the control group. Furthermore, individuals with mLD spent significantly less time in the Paint program than in the E-FIT program. This pattern of performance was not observed for members of the control group. Third, participants with mLD requested to change significantly fewer facial features than members of the control group. Moreover, they also requested to see significantly less exemplars before expressing being happy with one compared to the control group. Fourth, the mLD group showed a higher tendency of acquiescence than members of the control group, by accepting on average significantly more presented exemplars than the control group. And finally,

significant associations were revealed between the IQ scores of the mLD group and the duration of the E-FIT construction and the overall amount of facial information provided during the CI.

Despite the differences between the mLD group and the non-LD group in performance during the composite construction, is it possible that the composites created by the mLD group were nevertheless good enough for an independent sample of participants to identify the target face on the basis of the composite images? This research question was addressed further during the composite evaluation phase.

8.3 Phase 2: Composite evaluation

During the second phase of this study, two different tasks were used to evaluate the quality of the obtained composites: a matching task and a likeness rating.

8.3.1 Matching task

Method

Participants: An independent sample of participants ($n = 46$; 23 – 65 yrs; $M = 41.51$ yrs; $SD = 12.61$; 23 males and 27 females) not involved in the prior composite construction phase took part in the evaluation phase. They were all students or staff drawn from the University of Abertay Dundee.

Design: A 2 x 2 mixed design was used, including one within-subject factor (group: mLD vs. controls) and one between-subject factor (description mode: photo vs. memory).

Materials: The matching task consisted of the 60 composites, created during the construction phase, and a 10 person line-up consisting of the five target faces and five

distractor faces. The distractor faces were randomly selected from a larger face sample. All faces in the line-up were Caucasian males with no distinctive features. The images were full-face poses with neutral expressions. Figure 8.4 shows an example of the matching task.

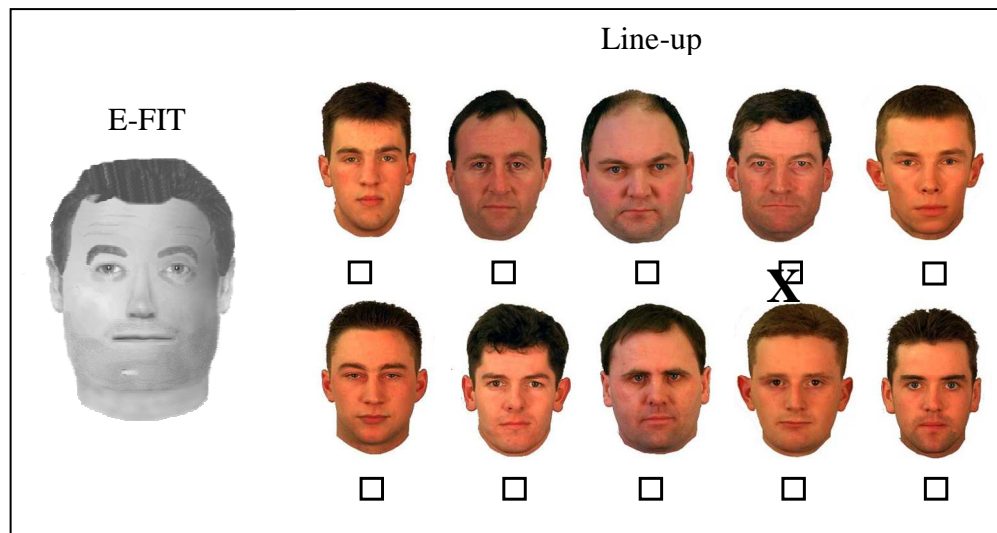


Figure 8.4. *Matching task: In this example, the participant has decided that the fourth face in the line-up (marked with an X) matches the composite on the left most.*

Procedure: The composites were presented to the participants in booklets. Each booklet contained all 30 composites obtained through the construction phase during one of the two description modes (either photo or memory condition). Each composite was displayed on its own page and accompanied by a 10 person line-up (see Figure 8.4). The order in which the composites were presented in the booklets was randomized as was the order in which the distractors and targets were presented in the line-ups. During the matching task, participants were asked to indicate which of the faces in the line-up best matched the accompanying composite.

Results

The average number of correct matches participants made on the basis of composites created by mLD participants and control participants was calculated, and are shown in Figure 8.5.

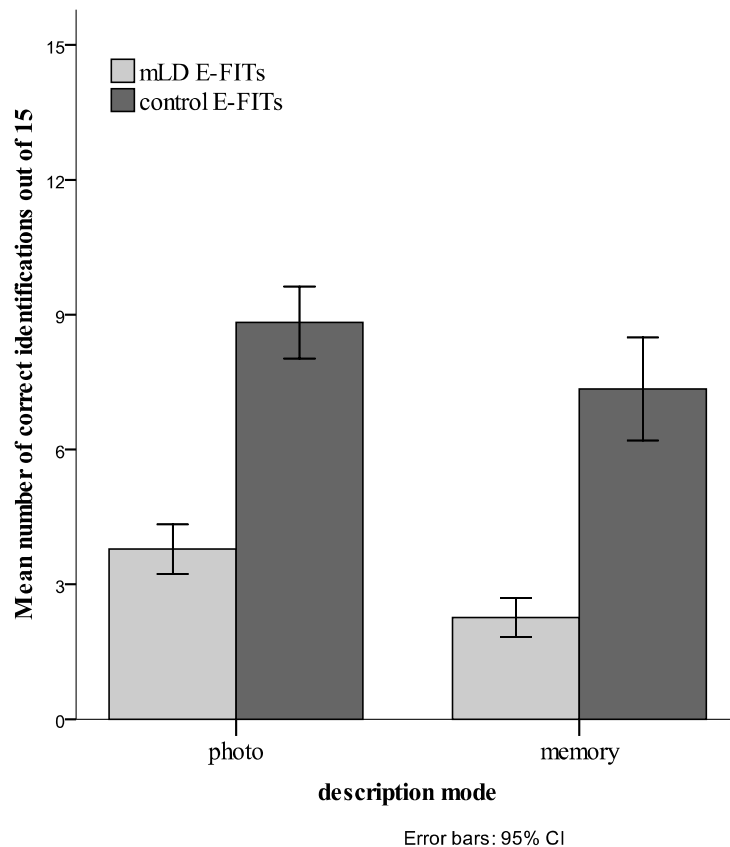


Figure 8.5. Mean number of correct matches (+SE) for E-FITs created by mLD and control participants during the different description modes.

A 2 x 2 mixed design ANOVA revealed a significant main effect for group $F(1, 44) = 181.07, p < .001, \eta^2 = .80$. Thus participants made significantly more correct matches when the composites were created by the control group than when they were created by the mLD group. Furthermore, a significant main effect for description mode was observed $F(1, 44) = 15.65, p < .001, \eta^2 = .26$, with participants making significantly more correct matches when the composites were created in the photo condition than in the memory condition. The interaction effect between group

and description mode was not significant $F(1, 44) = .003, p = .954, \eta^2 = .00$. See Figure 8.5 for means.

To further explore whether participants made more correct matches on the basis of facial composites created by mLD individuals than would have been expected by chance alone, the critical region of the binomial distribution was calculated. The binomial distribution had a mean of $pn = (1/10)(30) = 3$. Therefore, the critical region for the whole group of participants was calculated. The binomial distribution had a mean of $pn = (1/10)(46 \times 30) = 138$ and a standard deviation of $\sqrt{npq} = \sqrt{1380(1/10)(9/10)} = 11.14$. To be significantly different from chance, the group score must be above (or below) the mean by at least $1.96(11.14) = 21.83$. Thus, with a mean of 138, the group would need to score above 159.83 ($138 + 21.83$) or below 116.17 ($138 - 21.83$) to be significantly different from chance performance. The group had a total score of 139 correct matches for composites created by mLD participants, which can be regarded as chance performance.

8.3.2 Likeness rating task

Method

Participants: Forty-six participants not involved in the previous composite construction phase, or the matching task, engaged in the likeness rating task. They were all students or staff from the University of Abertay Dundee (27 – 65 yrs; $M = 42.27$ yrs; $SD = 11.28$; 14 males and 32 females).

Design: A 2 x 2 mixed design was employed, including one within-subject factor (group: mLD vs. controls) and one between-subject factor (description mode: photo vs. memory).

Procedure: The likeness rating task consisted of booklets including the 60 composites, created during the construction phase, alongside with the corresponding target faces. Participants were asked to indicate how good each composite was by rating how well it resembled the target on a scale ranging from no similar likeness at all (0) to very similar likeness (10). See Figure 8.6 for an example of the likeness rating task.

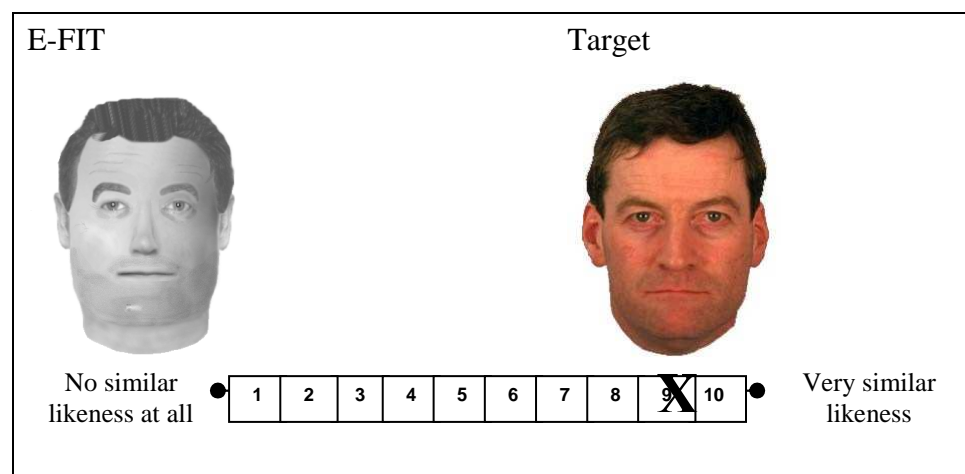


Figure 8.6. *Likeness rating task: In the example, the participant has decided that the composite is a quite similar likeness of the target; the participant has provided the composite with a score of 9 (marked with an X).*

Results

The average likeness rating scores participants gave to composites created by the mLD group and the control group during the two different description modes (photo vs. memory) were calculated and are shown in Figure 8.7. The likeness rating scales ranged from 1 to 10. A 2 x 2 mixed design ANOVA revealed a significant main effect for group $F(1, 44) = 340.04, p < .001, \eta^2 = .88$. Thus, participants rated the composites constructed by the control group as significantly more similar to the target face as the composites created by the mLD group. The main effect of description mode $F(1, 44) = .135, p = .715, \eta^2 = .00$ and the interaction between group and description mode $F(1, 44) = .715, p = .403, \eta^2 = .00$ were not significant.

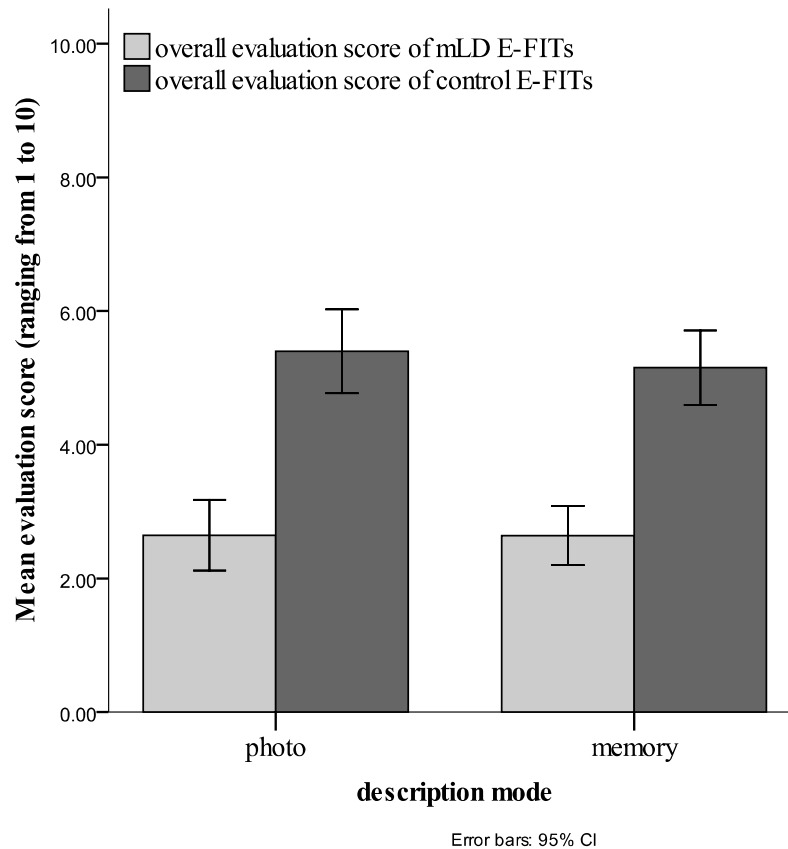


Figure 8.7. Mean likeness rating scores (+SE) for composites created by mLD and control participants during the different description modes (photo vs. memory).

8.3.3 Factors influencing composite quality

To investigate whether WASI scores, the duration of the composite construction phase, the overall amount of facial information obtained during the CIs, participants' non-verbal behaviours (such as pointing to their own face and pointing to the monitor), the number of features changed, and the number of exemplars presented had an influence on the quality of the resulting composites, the data obtained during the matching task and the likeness rating task were collapsed for description mode and scored in an additional way, which is outlined in more detail underneath.

Matching task

The overall amount of correct matches for each composite created by either an mLD participant or a control participant during the different description modes was calculated. Therefore, every participant, of the 60 participants originally taking part in the composite construction phase, received an overall matching score. For example, if 20 participants of the 46, who engaged in the matching task, matched a specific composite correctly, the participant who created this specific composite would receive an overall matching score of 20. The matching task scores of the mLD and the control group were then separately correlated with the duration of the composite construction phase, the overall amount of facial information obtained during the CIs, participants' non-verbal behaviours (such as pointing to their own face and pointing to the monitor), the number of features requested to be changed, and the number of exemplars presented to them.

The statistical analysis did not reveal any significant correlations for the mLD participants (all p s > .05). For the control group there was a significant positive correlation obtained between the number of correct matches and the duration of the composite construction phase ($r = .56$, $p = .001$). Moreover, the non-verbal behaviours of the control group, such as pointing to their own face to facilitate witness-operator communication ($r = .38$, $p = .04$) and pointing to the monitor ($r = .49$, $p = .006$), were significantly correlated with the number of correct matches obtained.

Likeness rating

The mean likeness rating score (ranging from 1 to 10) for each of the 60 original participants, involved in the composite construction phase, was calculated by adding

up the likeness scores provided by the 23 participants during the evaluation phase. The data were collapsed for description mode. The correlational analysis did not produce any significant associations for the mLD group (all $ps > .05$). For the control group, a significant positive correlation was obtained between the mean likeness rating score and the number of facial information provided during the CI ($r = .39, p = .03$). Furthermore, the mean likeness rating score was significantly correlated with the number of times control participants pointed to the monitor during the E-FIT construction process ($r = .42, p = .02$).

8.3.4 Summary of main findings

During the evaluation phase, the quality of the resulting composites was examined. The findings of the matching task and the likeness rating task showed that composites created by control participants were much more accurate than those created by the mLD participants. Specifically, it was found that participants made significantly more correct matches on the basis of composites created by the control group than by the mLD group. There were no associations obtained between the IQ scores of participants with mLD and the quality of the resulting composite images. Nor were there any significant correlations obtained between mLD composites' quality and any of the other measured variables, such as non-verbal behaviours, facial information provided during the CI, duration of the composite construction phase or features requested to be changed during the composite construction. However, the quality of composites created by the control group was significantly correlated with the duration of the composite construction phase, the information obtained during the CIs and the non-verbal behaviours provided during the composite constructions.

8.4 Discussion

In this study the ability of people with mLD to create facial composites with the E-FIT system was examined. The findings of the evaluation tasks showed that composites created by control participants were much more accurate than those created by individuals with mLD. The difference in the quality of the resulting composites might be due to several factors. First of all, during the CIs, participants with mLD reported overall less facial information than control participants. This research finding is consistent with previous research, showing that the accounts of people with mLD about an observed event are accurate but incomplete (Agnew & Powell, 2004; Milne, 1999). The incompleteness of the initial facial descriptions might have influenced the eventual quality of the resulting composites, since they determine the starting point of the composite construction process. For the control group a significant association was revealed between the amount of facial information reported during the CIs and the resulting composite quality. No such correlations were obtained for members of the mLD group. Thus, it seems that the more information control participants provide during the CIs the superior the resultant composite images will get. However, this rule of thumb does not hold for individuals with mLD.

Furthermore, individuals with mLD not only provided less verbal information during the CIs, they also exhibited less non-verbal information, such as pointing to their own faces or to the monitor to describe or to emphasise the size or position of individual facial features. It seems likely that such behaviours would facilitate witness-operator communication and assist in the accurate construction of the facial composite images. Control participants used significantly more non-verbal information to describe the target faces which might have led, among other factors, to superior quality facial composites. This conclusion was confirmed by the finding that

the amount of non-verbal behaviours provided by control participants was significantly associated with the quality of the resulting composites, thus the more non-verbal behaviours control participants used to describe the target faces, the better the resulting composite images were.

Secondly, participants with mLD spent less time constructing the composites than control participants. This was likely due to the fact that mLD participants requested fewer changes to be made to the composites and that they were much more easily satisfied with the resulting composites than control participants. It should be kept in mind here, that the composite construction process was completed when the participant stated that he/she was satisfied with the likeness of the resulting composite. This assumption is supported at least for members of the control group; the duration of the composite constructions was significantly correlated with the quality of the resulting composite images. No such association was revealed for the mLD group.

For the mLD group a significant positive correlation was found between IQ and the duration of the E-FIT construction phase, indicating that mLD participants with higher IQs spent more time engaging with the operator and the E-FIT software than mLD individuals with lower IQs. Participants with mLD also spent significantly less time amending the composites further with the aid of the Paint program compared to participants without LD. This may have had a detrimental effect on the resulting quality of the composite image, since research has shown that the quality of composites can be significantly improved by the use of graphic packages (Gibling & Bennett, 1994).

Earlier research demonstrated that people with LD have a higher tendency to acquiesce (Gudjonsson & Henry, 2003; Henry & Gudjonsson, 1999; Henry &

Gudjonsson, 2003; Milne, Clare & Bull, 2002). Similarly, a high tendency of acquiescence in the mLD sample was also observed in this study. Participants with mLD requested to see significantly fewer feature exemplars before selecting one which they regarded as a similar likeness than control participants. This shows that control participants are more fastidious and precise in their selections. Moreover, participants with mLD in general accepted more feature exemplars than members of the control group, which again confirms that they are more likely to respond in a positive way to presented stimuli and are more likely to please the operator.

With regard to the intellectual functioning of people with mLD, no significant associations were obtained during the present study, and the quality of the created facial composites. Thus, using psychometric IQ tests to determine an individuals' composite construction abilities does not seem to be an appropriate way of assessment and should not be used as justification for excluding individuals with mLD from general police procedures.

Overall, it can be concluded from the findings obtained during Study 3, that the E-FIT system is perhaps not the most suitable facial composite system for witnesses with mLD. The following studies investigate ways that might assist participants with mLD during the construction of facial composites. The effectiveness of visual prompts as facilitating tool and the suitability of an evolutionary facial composite system, namely EvoFIT, are examined during the subsequent studies.

Chapter 9

Study 4: Do visual prompts facilitate verbal descriptions of unfamiliar faces in witnesses with mLD?

This Chapter describes Study 4, which investigates the facilitating effect of visual prompts on the verbal descriptions of unfamiliar faces by people with mLD. Research in the past with children has shown that visual as well as verbal cues can enhance children's recall in quantity and quality (Aschermann, Dannenberg & Schulz, 1998; Paine, 2004). The study described in this chapter examines whether a similar effect is found when visual prompts are made available to adults with mLD during a facial description task. Participants were asked to describe unfamiliar faces with the target face in view or from their memory alone with the aid of visual prompts.

9.1 Introduction

Studies 2 and 3 revealed that verbal descriptions of faces are somewhat limited in people with mLD. This finding is in agreement with the research literature on event recall by people with LD, which has repeatedly found that their recall is poorer in quantity compared to the recall provided by people without LD (Agnew & Powell, 2004; Dent, 1986; Henry & Gudjonsson, 1999; Michel et al., 2000; Perlman et al., 1994). However, Study 2 also revealed that there may be circumstances from which individuals with mLD benefit when describing unfamiliar faces, such as a combination of open and more specific questions and describing the target face with the photograph in view compared to from memory alone. The open-questioning

approach elicited the most accurate information, whereas the cued-questioning approach had a beneficial impact on the completeness of the facial information provided. Similar results have been reported by Perlman et al. (1994). They demonstrated that a combination of free recall and specific questions elicited the most complete and accurate accounts about a to-be-remembered event from individuals with mLD.

Previous studies carried out with children have shown that the use of prompts can help them to remember more information. This was true for memories about a to-be-remembered event (Aschermann et al., 1998) and facial stimuli (Paine, 2004) (see Pipe, Lamb, Orbach & Esplin, 2004 for a literature review on the impact of cues and prompts on children's event recall). In a study conducted by Ascherman et al. (1998) children engaged in an interactive event and were questioned 10 days later with three different retrieval procedures. Procedure 1 included no retrieval aids, procedure 2 contained photographs of relevant topics and procedure 3 consisted of photographs and training regarding how to use these retrieval aids. It was found that children reported significantly more accurate and fewer inaccurate event details when questioned with the two procedures that included retrieval aids. Similar findings are reported in a study carried out by Paine (2004). She investigated the effect of prompts on children's verbal descriptions of unfamiliar faces. Children viewed faces of Caucasian males. After a 24-hour delay, participants were required to freely report what they could remember about the observed video sequences. The free recall was followed by either a verbal or visual prompt interview condition. During each prompt interview condition children were asked to select those prompts which most resembled the facial features of the target face. It was found that children provided significantly more accurate facial information during the prompted interview

conditions compared to the free recall condition. Furthermore, significantly more accurate facial information was obtained during the visual prompt condition than during the verbal prompt condition. Paine (2004) concluded that the use of prompts, particularly visual prompts, may enhance children's verbal descriptions of unfamiliar faces and might be therefore a useful tool in eyewitness situations, such as during the construction of facial composite images.

It could be argued that people with mLD display similar cognitive functioning as children, therefore may benefit from the same retrieval aids when verbally describing unfamiliar faces. Moreover, research has demonstrated that difficulties might arise during the composite construction process due to problems with the witness-operator communication (Brace et al., 2006). The use of visual prompts would decrease such difficulties, by minimising ambiguities and therefore making it easier for the operator to translate and interpret the verbal descriptions provided by the witness.

The present study required individuals with mLD to select visual prompts which resembled the facial features of two target faces most. As in the previous experimental studies, the control condition included people without LD. Two description modes were included: photo and memory. Before engaging in the actual experimental task, participants completed a practice task, during which they had to select prompts that most resembled the eyebrows of a female target. The purpose of the practice task was to familiarise participants with the experimental instructions and procedure.

9.1.1 Research aims

The aim of the current study was to investigate whether visual prompts are a useful tool to assist individuals with mLD when they have to verbally describe an unfamiliar

face, such as during the construction of facial composites with E-FIT. The hypothesis is that if individuals with mLD are able to use the visual prompts in an effective way there should be a considerably high level of agreement between the selected prompt options between the two groups of participants (mLDs and controls) as well as within each participant group.

9.2 Method

9.2.1 Design

A 2 (group: mLD vs. control) x 2 (description mode: memory vs. photo) mixed design was used, including one between subject factor (group) and one within-subject factor (description mode). The dependent variable was the amount of agreement between participants.

9.2.2 Participants

Overall, 43 participants took part in the study (19 males and 24 females). Of those, 22 were people with mLD (21 - 72 years; $M = 46.55$ yrs; $SD = 11.39$; WASI: FSIQ-4 score: $M = 59.14$, $SD = 4.91$; WASI: verbal score: $M = 56.62$, $SD = 3.07$; WASI: performance score: $M = 66.62$, $SD = 6.68$) and 20 were participants without LD (20 - 59 years; $M = 34.37$ yrs; $SD = 9.20$). One of the mLD participants did not complete the WASI and was therefore excluded from the dataset and the further analysis. Participants with mLD were recruited from day care centres in Dundee. They all had a WASI: FSIQ-4 score between 50 and 70, which fell within the desired range. Participants without LD were students and staff from the University of Abertay Dundee.

9.2.3 Materials

Psychometric Tests

As with previous studies, the WASI was used to assess verbal and non-verbal performance, as well as general intellectual functioning of participants with mLD.

Target faces

Target faces were full-face photographs of Caucasian males. There was a total of two targets. All targets were photographed without spectacles or other distinguishing marks. The facial expression of the targets was neutral and all targets were shown from the front. Each target photograph was 600 x 800 pixels in size. The targets derived from the same database as the ones used during Study 2. Figure 9.1 displays the two target faces.



Figure 9.1. *Target faces used during the visual prompt task (target face 1 is displayed on the left hand side and target face 2 is displayed on the right hand side).*

Visual prompts

The visual prompts were designed by Paine (2004) as part of her PhD studies with children (see Paine, 2004, pp. 185-188). Overall, there were 24 visual prompts including the seven main features of the face (eyebrows, eyes, nose, mouth, ears, hair

and face shape) and prompts regarding their size, colour and position (Figure 9.2 displays the visual prompt for eyebrow size). All visual prompts matched the E-FIT facial features in the Aberdeen Index. Each visual prompt was accompanied by a question mark, which provided participants with the possibility to answer “I don’t know” in the case they did not know or were not able to remember the appearance of the facial feature (see Figure 9.2). The complete set of visual prompts is depicted in Appendix 4.

1.2 Eyebrow Size

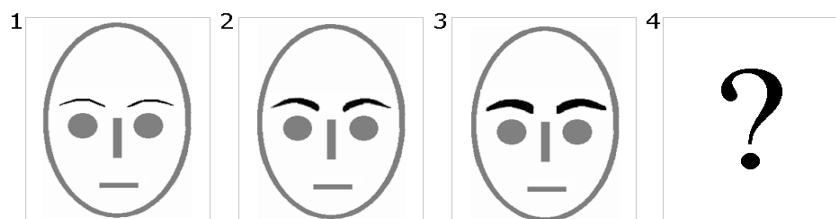


Figure 9.2. *The visual prompt for eyebrow size with the three different prompt options: thin, medium-sized or thick. Participants were also allowed to choose the “I don’t know option.”*

9.2.4 Procedure

A pilot study was carried out with three participants (including two with mLD and one without LD) to validate the experimental procedure and to ensure that participants with mLD understood the oral instructions. Participants were asked to repeat back the instructions given to them to demonstrate understanding. After having established that the three pilot participants were able to complete the task without any disruptions and without displaying any signs of distress, the actual experiment commenced.

At the beginning of each experimental session, written consent was obtained from the participants. The experimental procedure was the same for each participant group (mLDs and controls). Participants took part in the study individually. All

instructions were administered orally. Participants were encouraged to ask questions at any time. All participants first completed the practice task, which took approximately three minutes. Thereafter the main experimental tasks followed. Each individual first engaged in the memory condition and subsequently in the photo condition to avoid memory interference. In the memory condition participants viewed one out of two target faces, for 10 seconds which was presented on a monitor of a Toshiba laptop. The laptop was 36.2 cm x 26.8 cm x 3.9 cm in size and had a screen resolution of 1280 x 800. Next, participants completed the visual prompt task, during which, they were required to provide a non-verbal description of the previously seen target face with the assistance of visual prompts. The visual prompts were presented on the same monitor as the target face. The experimenter guided participants through the visual prompt task by providing verbal prompts in addition to the presented visual prompts (for example, while presenting the visual prompt for eyebrow size with the three different prompt options: thin, medium-sized or thick, the experimenter provided the verbal prompts: thin, normal or bushy). Participants were asked to select the visual prompts which resembled the facial features of the previously seen target most by pointing at them. The experimenter entered the participants' choices manually. All visual prompts were presented in a predetermined order which is consistent with the presentation order of the E-FIT facial descriptions during the construction of a facial composite based on the Aberdeen Index. Thereafter, the photo condition followed, during which the target face was visible all the time until the visual prompt task was completed. The overall experimental procedure was the same during the memory and the photo condition. During each condition, participants described one of the two faces. Thus, over the course of the experiment each participant described two faces. The presentation order of the faces was

counterbalanced in advance across participants. All participant responses were saved for later analysis. Participants were not provided with any feedback regarding the accuracy of their responses. At the end of the experiment, participants were fully debriefed and thanked for their participation. The whole experimental procedure lasted approximately 30 minutes.

9.2.5 Scoring

To analyse the data, a scoring sheet was devised, based on the most frequently selected prompt option by control participants. Tables 9.2 and 9.3 show which options were selected most often along with the percentage of agreement between control participants. The responses of each mLD and control participant during both description modes were then coded in accordance with this scoring sheet. For example, a participant who selected 12 visual prompt options that were consistent with the scoring sheet would receive a consistency score of 12 (out of 24). This scoring scheme made it possible to compare the performance of individuals with mLD with that of people without LD during the two different description modes. The data were collapsed across target face.

9.3 Results

The data analysis focused on the following research questions: First, is there a significant difference between the two participant groups? Second, is there a difference in the consistency scores during the two description modes? With regard to the first research question, no directional hypothesis was put forward, since the primary objective of this chapter was to explore whether visual prompts could be used by mLD individuals in a similar way as non-LD individuals would do to accurately describe unfamiliar facial stimuli. For the second research question, it was assumed

that mLD and control participants would select more consistent prompt options during the photo than the memory condition, since the memory condition can be regarded as the more cognitively demanding one.

A 2 (group: mLDs vs. controls) x 2 (description mode: photo vs. memory) mixed ANOVA was carried out. The between-subject factor was group and the within subject-factor description mode. The dependent variables were the memory consistency score and the photo consistency score for mLD and control participants. The Levene's test was significant for one of the dependent variables (photo consistency score: $p = 0.23$) and marginally significant for the other one (memory consistency score: $p = 0.54$), therefore the data were log transformed (\ln) and reanalysed. However, the Levene's test remained significant (photo consistency score: $p = .001$ & memory consistency score: $p = .001$). As a result, the data were analysed with the Mann-Whitney test. For the photo and memory consistency scores significant differences between mLD and control participants were revealed (photo: mLDs: $Mdn = 9$, controls: $Mdn = 18.5$; memory: mLDs: $Mdn = 7.5$, controls: $Mdn = 13$) (photo consistency scores: $U = 12.50$, $p < .001$, $r = -.81$; memory consistency scores: $U = 47.00$, $p < .001$, $r = -.67$). Thus, during both description modes (photo & memory), mLD individuals selected significantly fewer visual prompt options which were consistent with the scoring sheet than members of the control group (see Table 9.1 for mean).

Table 9.1. Mean number of prompt options (out of 24) (+SD) selected by mLD and control participants which were consistent with the scoring sheet.

Group	Number of consistent prompt options			
	Photo condition		Memory condition	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
mLDs	10.27	3.87	7.50	3.71
Controls	18.60	2.44	13.10	2.36

To explore further the impact of description mode on participants' selection, the data were split for group and analysed with two separate Wilcoxon signed-ranks tests. For the mLD group, the test revealed that they chose significantly fewer prompt options that were consistent with the scoring sheet during the memory condition compared to the photo one ($T = 42$, $p = .018$, $r = -.50$). A similar result was obtained for members of the control group ($T = 0$, $p < .001$, $r = -.82$). Thus, in line with the previously stated hypothesis, both participant groups selected significantly more prompt options that were consistent with the scoring sheet during the photo condition than during the memory condition (see Table 9.1).

9.3.1 Analysis of data with a more stringent cut-off criterion

As mentioned earlier in the scoring section of this chapter, the prompt options included in the scoring sheet were those most frequently selected by the control group. It could be argued that the applied cut-off criterion used to determine which prompt options are included in the scoring sheet was too lenient. Therefore, the scoring sheet was modified, including only those visual prompts which received an agreement level of 75% or above for one specific prompt option by members of the control group. As a result, the new scoring sheet included only 14 prompts instead of 48. The data were coded and analysed a second time with the more stringent cut-off criterion.

A 2 (group: mLDs vs. controls) x 2 (description mode: photo vs. memory) mixed ANOVA was conducted. As during the first analysis, the dependent variables were the memory consistency score and the photo consistency score for mLD and control participants. The Levene's test was significant for the photo consistency score ($p = .035$). Consequently, the data were log transformed (\ln) and reanalysed. The log transformation did not help to stabilise the data, the Levene's test remained significant

($p < .001$). Accordingly, the data were analysed with the Man-Whitney test. Similar to the finding obtained during the initial analysis, it was found that during both description modes (photo & memory), mLD participants selected significantly fewer prompt options which were consistent with the scoring sheet than control individuals (memory condition: $U = 61.50$, $r = -.62$, $p < .001$; photo condition: $U = 6$, $r = -.84$, $p < .001$).

Additionally, the data were split for group and analysed with two Wilcoxon signed-ranks tests to investigate the impact of description mode on selected prompt options. For the mLD group, the test showed no significant difference in the amount of selected prompt options which were consistent with the scoring sheet between the two description modes ($T = 51$, $p = 0.76$, $r = -.38$). Contrary, control participants selected significantly more prompt options which were consistent with the scoring sheet during the photo condition than during the memory one ($T = 11$, $p = .001$, $r = -.73$).

Thus, even when including only those visual prompts in the scoring sheet which received a considerable high level of agreement (75% or above), mLD individuals were nevertheless significantly less likely to select those options compared to their non-LD peers.

9.3.2 Correlations between performance and WASI scores

No significant correlations were obtained between the amount of consistent prompt options selected by mLD individuals and their verbal, performance or full-WASI scores (all $ps > .05$) (see Table 9.2 for the exact Spearman's rho correlation coefficients and p values).

Table 9.2. *Correlations revealed between mLD individuals' WASI scores and the amount of selected prompt options which were consistent with the scoring sheet (Spearman's rho correlation coefficients and p values are provided).*

	Number of consistent prompt options			
	Photo condition		Memory condition	
	<i>r</i>	<i>p</i>	<i>R</i>	<i>p</i>
Verbal WASI score	.317	.162	.016	.945
Performance WASI score	.368	.101	-.121	.602
Full WASI score	.383	.087	-.120	.603

9.3.3 In depth exploration of the raw data

When considering the raw data in more detail, it is noticeable that there was a high level of agreement for the most frequently selected prompt options amongst participants within the control group (see Table 9.3 and 9.4). For target face 1, participants only reached agreement of 45% and 40% for eye colour and face shape, respectively. However, for all other prompts (22) more than half of all participants agreed (> 50%) on one specific prompt option, even up to an agreement level of 95% (eyebrow colour).

A similar response pattern was obtained for target face 2, again for the majority of prompts (18 out of 24) over 50% of the participants agreed on one specific prompt option. For the prompt eyebrow shape and hair colour an agreement level of up to 95% was reached. For four prompts (eyebrow spacing, eye colour, hair style, eye & eyebrow spacing) half of the participants (50%) agreed on a distinct prompt option and only two prompts (ear setting and nose, mouth and chin spacing) received less than 50% agreement.

For target face 1, the highest agreement level received the prompt eyebrow colour (95%) followed by hair colour (90%). For target face 2, the prompts eyebrow

shape and hair colour (both 95%) revealed the highest amount of agreement, followed by eyebrow colour, eye shape, and mouth shape (all 80%). The lowest agreement levels were revealed for the prompts eye colour (40%), eye size, and face shape (both 45%) for target face 1, and ear setting (45%) and nose, mouth and chin spacing (45%) for target face 2.

Thus, regardless of target face, among control participants the prompts hair colour and eyebrow colour received the highest level of agreement ($\geq 90\%$) and eye colour and face shape the least ($\leq 50\%$).

Table 9.3. *The most frequently selected prompt options by control participants (n = 20) for target face 1 (collapsed across description modes) accompanied by the percentage of participants who selected them.*

	Eyebrow shape	%	Eyebrow size	%	Eyebrow colour	%	Eyebrow spacing	%
Answer	curved	65	thick	60	brown	95	average	55
	Eye shape	%	Eye size	%	Eye colour	%	Eye spacing	%
Answer	oval	65	small/average	45	blue	40	average	75
	Nose length	%	Nose breadth	%	Nose tip	%	Lips	%
Answer	short	75	wide	55	upturned	60	thick	55
	Mouth shape	%	Mouth width	%	Ear shape	%	Ear size	%
Answer	upturned	60	wide	55	rounded	85	average	80
	Ear setting	%	Hair length	%	Hair style	%	Hair type	%
Answer	close to head	65	short	75	parting	75	slightly wavy	65
	Hair colour	%	Face shape	%	Eye & Eyebrow spacing	%	Nose, mouth & chin spacing	%
Answer	brown	90	round	45	close	60	close	65

Table 9.4. *The most frequently selected prompt options for the different facial features by control participants (n = 20) for target face 2 (collapsed across description modes) accompanied by the percentage of participants who selected them.*

	Eyebrow shape	%	Eyebrow size	%	Eyebrow colour	%	Eyebrow spacing	%
Answer	straight	95	thick	60	black	80	average	50
	Eye shape	%	Eye size	%	Eye colour	%	Eye spacing	%
Answer	oval	80	large	50	blue	55	average	70
	Nose length	%	Nose breadth	%	Nose tip	%	Lips	%
Answer	average	65	narrow	55	straight	55	Thin	70
	Mouth shape	%	Mouth width	%	Ear shape	%	Ear size	%
Answer	straight	80	average	75	pointing	65	average	60
	Ear setting	%	Hair length	%	Hair style	%	Hair type	%
Answer	protruding	45	short	50	spiky	55	straight	55
	Hair colour	%	Face shape	%	Eye & Eyebrow spacing	%	Nose, mouth & chin spacing	%
Answer	black	95	angular	50	close	55	average	45

In contrast to control participants, individuals with mLD showed considerable less agreement for most of the prompts (see Tables 9.5 and 9.6). For target face 1, only 4 (out of 24) prompts received a majority vote (> 50%) for a distinct prompt option. For target 2, only 6 (out of 24) prompts received a majority vote (> 50%). Thus, regardless of the target face presented, for the majority of visual prompts participants with mLD were not able to agree up to a considerable amount (> 50%) on one specific prompt option. For target face 1, the highest level of agreement received the prompt hair colour (64%); followed by eyebrow shape, eye shape, and ear shape (all 55%). The lowest amount of agreement was reached for the prompts nose tip, lips, ear size, and hair length (36%). For target face 2, the prompt eye size received the highest agreement level (73%), followed by eyebrow shape (59%). The lowest agreement received the prompt lips (27%), followed by eye spacing, nose length, and hair style (all 32%). Overall, regardless of the target face viewed, participants with

mLD displayed the highest level of agreement for the visual prompts eyebrow shape and hair colour ($\geq 55\%$).

Table 9.5. *The most frequently selected prompt options by mLD participants (n = 22) for target face 1 (collapsed across description modes) accompanied by the percentage (rounded) of participants who selected them.*

	Eyebrow shape	%	Eyebrow size	%	Eyebrow colour	%	Eyebrow spacing	%
Answer	curved	55	thin	46	brown	46	average	50
	Eye shape	%	Eye size	%	Eye colour	%	Eye spacing	%
Answer	oval	55	average/large	41	blue	41	average	41
	Nose length	%	Nose breadth	%	Nose tip	%	Lips	%
Answer	average	46	average	41	straight/ downturned	36	very thick	36
	Mouth shape	%	Mouth width	%	Ear shape	%	Ear size	%
Answer	upturned	50	wide	46	rounded	55	average	36
	Ear setting	%	Hair length	%	Hair style	%	Hair type	%
Answer	average	46	very short	36	brushed forward	41	straight	41
	Hair colour	%	Face shape	%	Eye & Eyebrow spacing	%	Nose, mouth & chin spacing	%
Answer	brown	64	round	46	apart	46	average	41

Table 9.6. *The most frequently selected prompt options by mLD participants (n = 22) for target face 2 (collapsed across description modes) accompanied by the percentage (rounded) of participants who selected them.*

	Eyebrow shape	%	Eyebrow size	%	Eyebrow colour	%	Eyebrow spacing	%
Answer	straight	59	thick	50	brown	50	apart	41
	Eye shape	%	Eye size	%	Eye colour	%	Eye spacing	%
Answer	oval	55	average/large	73	blue	55	average	32
	Nose length	%	Nose breadth	%	Nose tip	%	Lips	%
Answer	average	32	wide	55	downturned	36	thick	27
	Mouth shape	%	Mouth width	%	Ear shape	%	Ear size	%
Answer	upturned	41	wide	46	rounded	50	average	46
	Ear setting	%	Hair length	%	Hair style	%	Hair type	%
Answer	protruding	46	very short	50	spiky	32	straight	46
	Hair colour	%	Face shape	%	Eye & Eyebrow spacing	%	Nose, mouth & chin spacing	%
Answer	black	55	round	46	average	41	average	36

Furthermore, as already established with the statistical analysis, the in depth exploration of the raw data emphasises that the most frequently chosen prompt options by participants with mLD were often not consistent with those selected most frequently by control participants. For target face 1, only 13 (out of 24) of the prompt options most frequently picked by mLD participants were consistent with those most frequently selected by control participants. The same finding was obtained for target face 2, again merely 13 (out of 24) of the prompt options most frequently picked by participants with mLD were consistent with those most frequently selected by members of the control group (Table 9.7 depicts the 13 prompts which received a consistent response between mLD participants and members of the control group for target face 1 and 2, respectively). Interestingly, the visual prompts for eyebrow shape and hair colour received considerable high levels of agreement by both participants groups and across the two target faces.

Table 9.7. *Display of prompts, for which the selected prompt option was consistent between mLD and control participants for target faces 1 & 2, respectively, together with the level of agreement in percentages within each participant group.*

Prompt	Face 1 mLD	Face 1 control	Prompt	Face 2 mLD	Face 2 control
Eyebrow shape	55%	65%	Eyebrow shape	59%	95%
Eyebrow colour	46%	95%	Eyebrow size	50%	60%
Eyebrow spacing	50%	55%	Eye shape	55%	80%
Eye shape	65%	55%	Eye size	73%	50%
Eye size	41%	45%	Eye colour	55%	55%
Eye colour	41%	40%	Eye spacing	32%	70%
Eye spacing	41%	75%	Nose length	32%	65%
Mouth shape	50%	60%	Ear size	46%	60%
Mouth width	46%	55%	Ear setting	46%	45%
Ear shape	55%	85%	Hair style	32%	55%
Ear size	36%	80%	Hair type	46%	55%
Hair colour	64%	90%	Hair colour	55%	95%
Face shape	46%	45%	Nose, mouth & chin spacing	36%	45%

Overall, the responses provided by mLD participants were much more varied than those given by control individuals, as evidenced by the low percentages of agreement in the mLD group. This was even the case for prompt options, which were most frequently selected by both participant groups. To examine whether this difference in agreement level between the two participants groups was significant, *t*-tests were carried out (the data was collapsed across description mode). The percentage of agreement served as dependent variable. For both target faces, the analysis revealed a significant difference between the mLD and control group in the levels of agreement (target face 1: $t(10) = -2.87, p = .017, r = .67$; target face 2: $t(10) = -2.56, p = .028, r = .62$). On average, for face 1, 50.82% (SD = 2.47) of the mLD participants agreed on a prompt option, whereas 65.90% (SD = 5.34) of control participants agreed. For target face 2 a similar pattern emerged, the agreement level for mLD participants was 46.55% (SD = 3.90) and 60.91% (SD = 4.56) for members of the control group.

9.4 Discussion

This study aimed to investigate whether people with mLD can use visual prompts effectively to describe unfamiliar faces. To approach this aim, it was determined at the beginning of the analysis which visual prompt options were selected most frequently by the control group. The responses by mLD and control participants were then scored in accordance with a scoring template that listed the most frequently selected options for each of the 24 visual prompts by control participants. The data analysis revealed that members of the control group selected significantly more often prompt options which were consistent with the scoring sheet than members of the mLD group. Thus, people with mLD did not choose those prompt options which were

selected most frequently by the majority of individuals without LD. Furthermore, it was revealed that both, mLD and control participants, selected significantly more prompt options which were consistent with the appropriate ones during the photo condition than during the memory one. This finding is in agreement with the previously stated hypothesis and confirms that the memory condition can be considered as the more cognitively demanding one. However, it should be noted that this pattern changed slightly when a more stringent cut-off criterion was employed. Although members of the control group still selected significantly more prompt options which were consistent with the scoring sheet during the photo condition than during the memory one; description mode had only a marginally effect on the selection by mLD individuals.

Unfortunately, the findings of the present study do not support the notion that visual prompts are a useful facial description aid for individuals with mLD. In contrast to participants without LD, individuals with mLD seemed to select the visual prompt options more randomly with no logical response pattern being recognisable. Contrary to control participants, participants with mLD displayed much more variability in their choices regarding which the appropriate prompt option was. Therefore it was felt that the utility of these visual prompts to facilitate or enhance the verbal facial descriptions of individuals with mLD was questionable. Consequently, the idea to use the suggested visual prompts during a subsequent facial composite construction study was set aside. Instead it was thought to be more fruitful to investigate composite systems which do not require a verbal facial description, such as EvoFIT.

Chapter 10

Study 5: The suitability of EvoFIT for witnesses with mLD

Holistic facial composite systems, such as EvoFIT do not require witnesses to provide a verbal description of a perpetrator's face but rely more on face recognition (Frowd et al., 2004; 2005b). As such they appear to be particularly suitable for witnesses with LD. This final empirical chapter describes Study 5, which investigates the suitability of EvoFIT for mLD witnesses. Two groups of participants, one with mLD and one without LD are required to use the EvoFIT system to create facial composites of unfamiliar faces. The quality of the resulting composites is assessed by an independent sample of participants via a matching task and a likeness rating task. The findings provide an indication of which facial composite system is more suitable for witnesses with mLD, E-FIT which can be regarded as a more featural system or EvoFIT which is more a holistic one.

10.1 Introduction

People find it very difficult to construct faces from memory (Frowd, Bruce & Hancock, 2008). Specifically, they have problems describing and selecting individual facial features. In part this may be due to the way we usually process faces. Research has shown that we process faces more as a whole rather than as a sum of its parts (see Maurer, Le Grand & Mondloch, 2002, for a review on configural face processing and Chapter 4 of this thesis for a description of research investigating featural and holistic face processing). As noted before, this might have serious consequences when it comes to the construction of facial composites. Current facial composite systems

utilised in the UK, such as ProFIT and E-FIT, do not encourage holistic face processing, instead the witness has to select individual facial features, such as hairstyle, eyebrows, nose, etc. (Frowd et al., 2005b). A detailed verbal description is therefore a prerequisite of facial composite construction (Frowd, Bruce & Hancock, 2008). This might be especially problematic if the witness or victim has LD. Prior research has revealed that individuals with LD often have limited verbal abilities (Emerson, 2001), which might act as a barrier to them providing a reliable description of a perpetrator's face. During Study 3, witnesses with mLD and without LD were required to use E-FIT to construct facial composites of unknown faces. The composites created by the mLD group were significantly poorer than those created by members of the control group. A subsequent analysis of the data revealed several contributing factors for the differing quality of the resulting composite images, such as participants with mLD providing significantly fewer verbal information about the target faces during the CI and being less critical and more easily satisfied with the resultant composite images than the control participants. A facial composite system has now been developed, which, in theory, should be more appropriate for people with LD. The innovative system is called EvoFIT and was designed by Professor Peter Hancock at the University of Stirling, Professor Vicki Bruce at Newcastle University and Dr Charlie Frowd at the University of Central Lancashire. According to its developers, it does not require a verbal description of the perpetrator's face but relies more on face recognition (Frowd et al., 2004; Frowd et al., 2005b). During the construction of facial composites with EvoFIT, witnesses select faces that look similar to the face of the perpetrator. The selected faces are then bred together to produce another set of faces. The selection and breeding of faces is repeated several times until a good likeness of the perpetrator's face emerges (Frowd et al., 2004).

The purpose of this experimental study is to investigate the ability of individuals with mLD and without LD to use the EvoFIT system. If the poor quality of the facial composites created with E-FIT in the previous study was due to mLD participants having problems with communication and language it would be hypothesised that their performance will not be significantly worse than that of witnesses without LD when using a system that relies less on verbal descriptions. If however, the poorer E-FIT composites were due more to problems with memory one would still hypothesise that mLD participants would generate poorer composites with EvoFIT compared to those created by members of the control group.

The study consists of three parts: the morph task, the composite construction and the composite evaluation. Given that different participants took part in each stage, each is described separately with its own method and result section.

10.2 Part 1: The morph task

The EvoFIT system requires witnesses to select faces which look similar to the perpetrator's face during the facial composite construction. It could be argued that the concept of similarity is quite an abstract one and research has shown that people with LD find it difficult to understand abstract language concepts (Bradshaw, 2001). Given this, a morph task was designed to examine whether individuals with mLD were able to understand the concept of similarity. Moreover, the task assessed their ability to differentiate between more and less similar looking faces with reference to a target face, an ability that has fundamental importance when it comes to the later composite construction with EvoFIT. During this task participants were presented with unfamiliar female faces and were required to make similarity judgments.

10.2.1 Method

Design

A 2 (group: mLD vs. control) x 2 (task: easy vs. difficult) mixed design was used. The dependent variable was the accuracy of the similarity judgments.

Participants

Forty participants took part in the morph task (16 males and 24 females). Of those, 20 were individuals with mLD (19 - 66 years; $M = 44.25$ yrs; $SD = 12.69$; WASI: FSIQ-4 score: $M = 59.70$, $SD = 4.74$; WASI: verbal score: $M = 59.15$, $SD = 4.65$; WASI: performance score: $M = 65.90$, $SD = 6.12$) recruited from day care centres in the Dundee area and 20 were individuals without LD, recruited from the student and staff body of the University of Abertay (20 - 61 years; $M = 41.80$ yrs; $SD = 11.62$). All individuals with mLD had a WASI: FSIQ-4 score between 50 and 70.

Psychometric tests

As during previous experiments verbal as well as non-verbal performance and general intellectual functioning of the mLD group was assessed with the WASI.

Materials

Twenty pairs of static full-face photographs of unfamiliar Caucasian females were used as stimuli. The dyads were selected from a larger sample of photographs provided by Professor Peter Hancock, Stirling University. From each dyad a morph image was generated with PsychoMorph (Tiddeman, Burt, & Perrett, 2001). The morph possessed characteristics of both original faces (see Figure 10.1 for an example. Note the morph is always presented above the two original faces, which

contributed to it). Half of the created morphs shared 70% of the characteristics of one face and 30% of the other face. Those triads formed the stimuli for the easy morph task. The other half shared 60% characteristics of one face and 40% of the other face. These constituted the stimuli for the difficult morph task. All stimuli were presented without spectacles or other distinguishing marks. The facial expressions were all neutral and faces were shown from the front. Each picture was 363 x 499 pixels in size. Each triad was displayed on a black background.



Figure 10.1. *Example of face triads used during the easy (left side) and difficult (right side) morph tasks. In the easy task, the morph in the top shares 70% characteristics of the face in the bottom left and 30% characteristics of the face in the bottom right. In the difficult task, the morph in the top shares 60% characteristics of the face in the bottom left and 40% characteristics of the face in the bottom right (the face in the bottom associated with a correct response is located randomly on the left and the right handside during presentation).*

Procedure

The experimental procedure was the same for both groups (mLDs and controls) and participants took part individually. All instructions were administered by the experimenter orally. During the morph task, participants were presented with 20 face triads via the Superlab software on a Toshiba laptop. The laptop was 36.2 cm x 26.8

cm x 3.9 cm in size and had a screen resolution of 1280 x 800. The morph task consisted of two parts: the easy task and the difficult task. Participants always first completed the easy morph task and then the difficult one. During each task the face triads were presented in a random order. For each triad, participants were asked to compare the two original faces with the morph and to decide which one of the originals looked more similar to the morph. Participants' responses were entered manually by the experimenter. Triads remained on the screen for as long as participants needed to make their decisions. The morph task lasted approximately 10 minutes.

Scoring

For both tasks the total amount of correct responses was calculated for each participant. A correct response was defined as selecting the original face which contributed to a higher degree to the morph, i.e. during the easy morph task, the correct response is the selection of the original face which contributed to 70% of the characteristics of the morph image. For the difficult task, the correct response is the selection of the original face which contributed to 60% of the characteristics of the morph. The location of the faces associated with a correct response was counterbalanced across task.

10.2.2 Results

The analysis of data focused on the following research questions: First, is there a significant difference between the performance of people with mLD and control participants during the easy morph task and the difficult morph task? Second, do people with mLD perform better than would be expected by chance alone? An alpha level of .05 was used for all statistical tests.

A 2 (group: mLD vs. controls) x 2 (task: easy vs. difficult) mixed design ANOVA was carried out. The Levene's test was significant (easy task: $p = .001$; difficult task: $p = .018$), hence the data was log transformed (ln) and reanalysed. The Levene's test remained significant (easy task: $p = .001$; difficult task: $p = .003$), therefore the file was split for task and analysed with the Mann-Whitney test. For both tasks a significant difference for the average number of hits was obtained between the mLD group and the control group (mLDs: easy task: $Mdn = 8.5$, difficult task: $Mdn = 7.5$; control: easy task: $Mdn = 10$, difficult task: $Mdn = 9$) (easy task: $U = 80.00$, $p = .001$, $r = -.55$; difficult task: $U = 105.50$, $p = .009$, $r = -.41$). Thus, mLD participants performed significantly poorer on both tasks than members of the control group (see Table 10.1 for means and standard deviations).

Table 10.1. Mean number of correct responses during the easy and difficult morph tasks (out of 10) (+SD) by mLD and control participants.

Group	Morph task			
	Easy		Difficult	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
mLDs	8.30	1.56	7.15	1.78
Controls	9.70	0.47	8.60	1.19

To further examine whether there were differences in performance for both groups between the difficult and easy morph task the data were split and analysed for each group (mLDs and controls) separately with two Wilcoxon signed-ranks tests. The analysis revealed a significant difference between the performance on the easy task and the difficult task for the mLD group ($T = 9.00$, $p = .006$, $r = -2.77$) as well as for the control group ($T = 10.00$, $p = .002$, $r = -3.07$). As expected, both groups performed better during the easy task ($\underline{M} = 9.00$, $\underline{SD} = 1.34$) than during the difficult task ($\underline{M} = 7.88$, $\underline{SD} = 1.67$).

To address the second research question, whether people with mLD performed better than would be expected by chance alone, the amount of correct responses on the easy and difficult morph task were collapsed, since both p_n and q_n must be at least 10 before one can use the binomial distribution to determine critical values for a binomial test. Next, it was calculated for each individual whether he/she scored significantly different from chance. The binomial distribution had a mean of $p_n = (1/2)(20) = 10$ and a standard deviation of $\sqrt{npq} = \sqrt{20(1/2)(1/2)} = 2.24$. To be significantly different from chance, a score must be above or below the mean by at least $1.96(2.24) = 4.39$. Thus, with a mean of 10, an individual would need to score above 14.39 ($10 + 4.39$) or below 5.61 ($10 - 4.39$) to be significantly different from chance. For the mLD group, 8 participants performed at chance and 12 significantly above chance level. All of the control participants performed above chance. A binomial test revealed that the proportion of mLD participants who performed above chance was not significantly different from those who performed at chance, $z = 0.91$, $p = .503$. Thus, mLD participants did not select the correct answers significantly more often than the incorrect ones.

When looking at the group performance of mLD individuals on both morph tasks (easy and difficult morph task) separately, the results look slightly different. For the whole group, the binomial distribution has a mean of $p_n = (1/2)(20 \times 10) = 100$ and a standard deviation of $\sqrt{200 \times (1/2)(1/2)} = 7.07$. To be significantly different from chance the group's score must be above or below the mean by at least $1.96(7.07) = 13.86$. Thus, with a mean of 100, the group would need to score above 113.86 ($100 + 13.86$) or below 86.14 ($100 - 13.86$) to be significantly different from chance. For the easy morph task, the group score was 166, which is significantly above chance. For the difficult morph task, it was 143, which although lower is still significantly above

chance. Thus, although some individuals with mLD scored at chance, as a group, they performed significantly better than expected by chance alone.

Finally, a correlational analysis was carried out on mLD participants' WASI scores and their number of correct responses during the morph tasks, to investigate whether there was an association between general intellectual functioning and participants' ability to make similarity judgments. No significant correlations were revealed (all $ps > .05$) (see Table 10.2).

Table 10.2. *Correlations obtained between mLD participants scores on the WASI and the morph tasks (Spearman's rho correlation coefficients and p values are provided).*

	Correct responses during the morph tasks					
	Easy task		Difficult task		Easy + Difficult tasks	
	<i>r</i>	<i>P</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>P</i>
WASI: FSIQ-4 score	.098	.680	-.088	.713	-.001	.996
WASI: verbal score	.008	.973	-.142	.549	-.081	.734
WASI: performance score	.108	.650	-.008	.973	.052	.829

Summary

Overall, the morph task demonstrated that individuals with mLD have more difficulties distinguishing more and less similar looking faces than individuals without LD. However, when considering group performance, the findings showed that the mLD group performed significantly better on the morph task than would have been expected by chance alone. Furthermore, although on a lower level, they displayed the same performance pattern as individuals without LD. They performed significantly better during the easy morph task than during the difficult one. On the basis of these findings we concluded that it seemed reasonable to investigate the suitability of the EvoFIT system with mLD individuals.

10.3 Part 2: Composite construction

Prior to the EvoFIT composite construction, participants, who did not take part during Part 1 of this Study, completed a practice task which consisted of the morph task. Thus, participants underwent the same morphing task as the participant group in Part 1 of this Study, using the identical method. The purpose of the practice task was to get participants used to the instructions by the experimenter and to examine whether there are any associations between performance on the morph task and the quality of the subsequent constructed composite images. The stimuli and the procedure during the practice task were the same as the ones applied during the morph task during Part 1 of this study. After the completion of the practice task, the composite construction followed. During the composite construction, participants produced facial composites together with the experimenter using the EvoFIT package. The experimenter received training in the use of EvoFIT by Professor Peter Hancock (Stirling University), who is one of the developers of this system.

10.3.1 Method

Design

A 2 (group: mLD vs. controls) x 2 (description mode: photo vs. memory) between-subjects design was used.

Participants

Overall, 64 participants took part in this study (29 males and 31 females). The sample consisted of participants with mLD and without LD. Participants with mLD were recruited from social day care centres in and around the Tayside area. Control participants were students and members of staff from the University of Abertay

Dundee. General intellectual functioning of participants with mLD was assessed using the WASI. Four participants with mLD had to be withdrawn from the data set because they had WASI scores above 70 and therefore did not classify as having mLD according to the WHO definitions (1992). After exclusion of the four participants the WASI: FSIQ-4 score ranged from 51 to 70 ($M = 58.10$, $SD = 6.24$) (WASI: verbal score: $M = 57.69$, $SD = 7.24$; WASI performance score: $M = 66.09$, $SD = 10.11$). There were 60 participants left in the final data set, of those, 30 were individuals with mLD (18-55 years, $M = 35.97$ yrs, $SD = 9.84$) and 30 were individuals without LD (19 - 38 years; $M = 24.33$ yrs; $SD = 4.74$).

Materials

Facial stimuli: Static full-face photographs of unfamiliar Caucasian males were used as stimuli. There was a total of 5 targets. All faces were photographed without any distinguishing marks.

Psychometric Tests: As in previous experiments, verbal as well as non-verbal performance and general intellectual functioning of participants with mLD was assessed with the WASI.

Facial composite system: The EvoFIT program was used to create facial composites. It ran on a Toshiba Satellite Pro A200 laptop running Windows XP.

Procedure

Prior to the actual composite construction, the experimenter tested the feasibility of the stimuli and the general experimental procedure with 2 participants (one with mLD and one without LD). Afterwards these two video-taped sessions were assessed by the experimenter together with her supervisory team. The assessment revealed that both

participants appeared to understand the oral instructions delivered by the experimenter, as they were able to repeat back what was requested from them. Moreover, both participants completed the experiment without any disruptions or signs of distress. On the basis of these findings the experimenter decided together with her supervisory team that the procedure and the stimuli were appropriate.

The experiment took place in a quiet room and participants took part individually. At the beginning of each experimental session, written consent was obtained from the participants. Given the possible influence of practice on a participant's behaviour, efforts were made to ensure that all participants had not previously constructed a facial composite. Next, participants with mLD completed the WASI. In line with previous experimental studies, participants created composites during two different description modes: photo vs. memory. In the photo condition the composite was created with the picture of the target face in view and participants were allowed to look as often and as long at it as they needed while constructing the composite image. In the memory condition the photograph was not present when the composite was created; instead participants viewed the target face for one minute prior to the actual composite construction. As in Study 3, the memory condition was divided into two sessions, a morning and an afternoon session. This created a delay of at least three hours between the actual presentation of the target face (morning session) and the construction of the composite (afternoon session). All participants selected the target faces randomly out of an envelope and the identity of the target was kept a secret from the experimenter during the whole composite construction process.

The procedure used to construct the composites followed the one outlined in the EvoFIT manual and as recommended by Professor Peter Hancock. Participants in

the memory condition received additional mental context reinstatement instructions prior to the composite construction. All participants first engaged in the selection and breeding phase of the EvoFIT system. This phase started with the selection of an appropriate hairstyle which resembled the one of the target face most. Thereafter participants were required to select a predetermined number of faces (24 overall) which looked similar to the target face (see Figure 10.2 for a screenshot of the selection and breeding phase). The selected faces were then bred together and a new generation of faces evolved. The process of selecting and breeding was repeated two times.

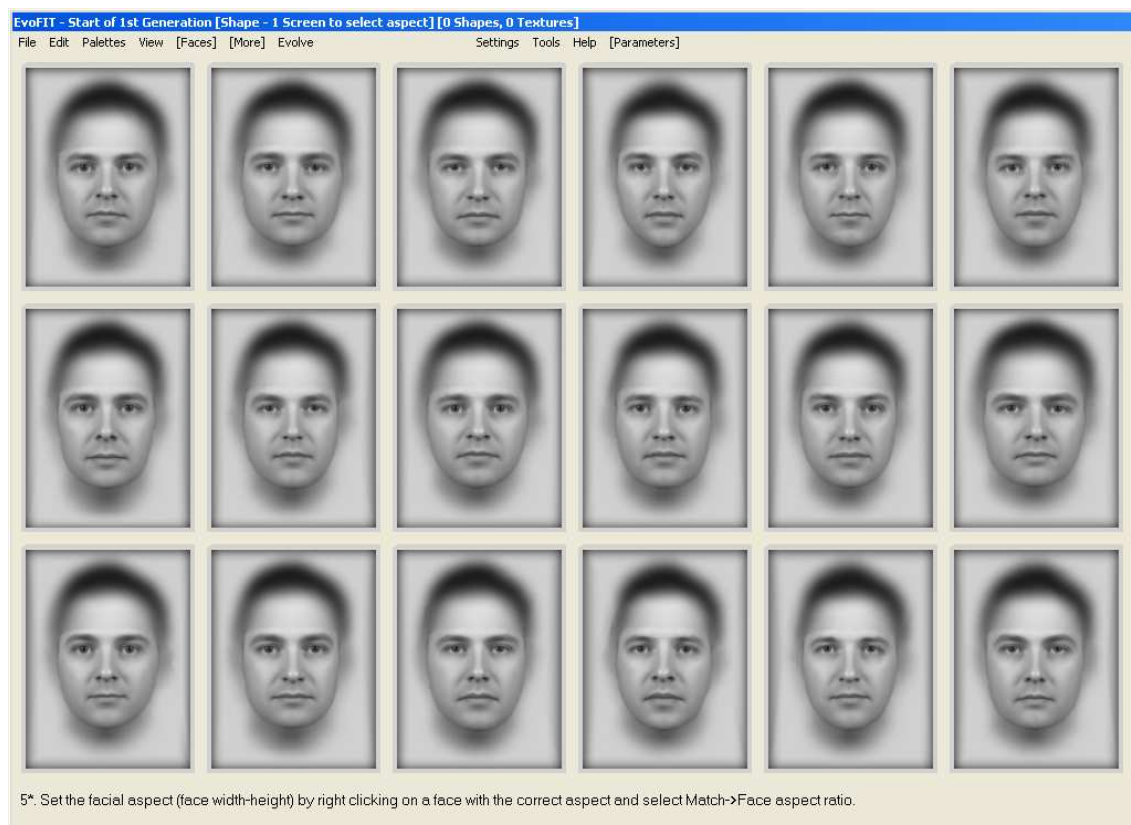


Figure 10.2. Screenshot of the selection and breeding phase during the composite construction procedure with EvoFIT.

Thereafter, the composite image was further edited with the help of the holistic tool. This tool changed the face on the request of the participants further by modifying for

example its age, width, masculinity or honesty (see Figure 10.3 for a screenshot of the holistic tool).

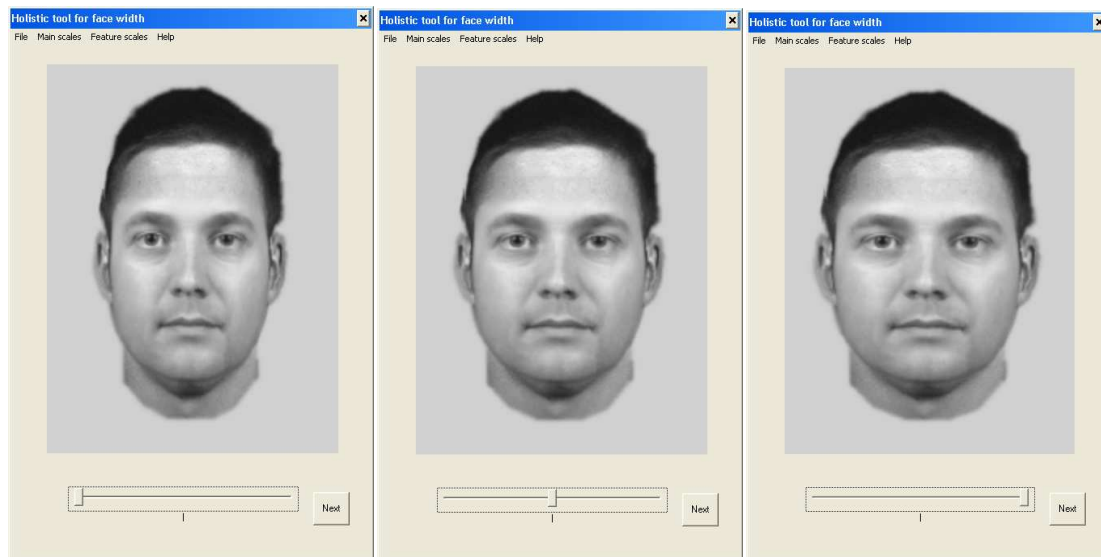


Figure 10.3. Screenshot of the holistic tool during the composite construction procedure with EvoFIT. The below presented holistic scale allowed participants to change the width of the face.

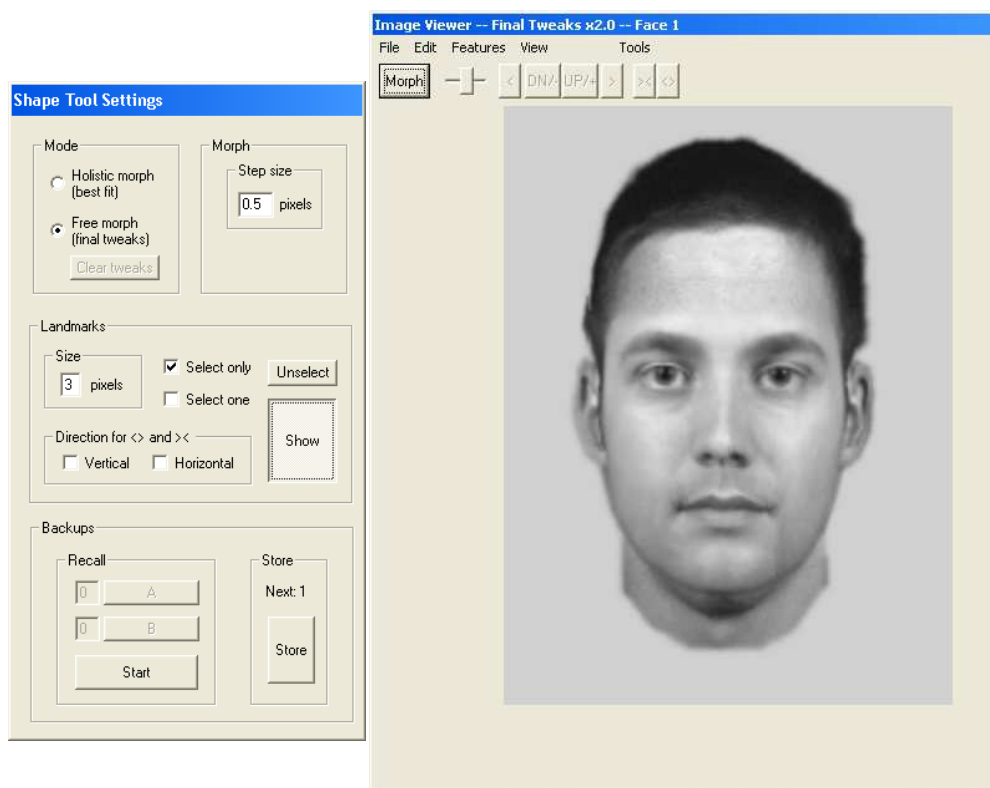


Figure 10.4. Screenshot of the shape tool during the composite construction procedure with EvoFIT.

Finally the size and/or position of individual facial features were changed with the aid of the shape tool (see Figure 10.4 for a screenshot of the shape tool). The experimenter finished the composite construction process when the participant stated that he/she was satisfied with the likeness of the resulting composite. Finally, participants were fully debriefed and thanked for their participation.

Scoring

The mean length of time participants spent in each of the three composite construction phases (selection and breeding, holistic tool and shape tool) was calculated. Moreover, it was examined whether participants accepted the changed face after having used the holistic tool or not. Additionally, the total amount of facial features (out of 10: cheeks, ears, eyebrows, eyes, face width, forehead, jaw/chin/jowls, mouth/lips, nose/nostrils/temp and philtrum) requested to be modified with the shape tool was counted for each participant. Correlational analysis showed a significant level of agreement for the total amount of changed facial features between two independent coders based on a random sample of 10 composite construction sessions ($r = .936, p < .001$).

10.3.2 Results

Research Questions

The data analysis concentrated on the following research questions. First, was there a significant difference between the two participant groups in performance on the practice task? Based on the findings revealed during the morph task, a significant difference was expected. Second, was there a difference in the duration of time spent by mLD and control participants in the three different composite construction phases

with EvoFIT (selection and breeding, holistic tool and shape tool) during the different description modes? Study 3 revealed that mLD individuals spent significantly less time with the construction of the composites with E-FIT than control participants. Therefore it was hypothesised that participants with mLD would spend less time than control participants during each phase. Third, do participants with mLD choose the changed composite after the application of the holistic tool as often as participants from the control group? Having established in Study 3 that participants with mLD are less critical, it was assumed that there would be a significant difference between the groups. Finally, are there differences between the two participant groups in the amount of facial features they want to change during the shape tool? Study 3 showed that the mLD group requested to change significantly fewer facial features than members of the control group during the composite construction with E-FIT, therefore it was hypothesised that there will be significant group differences.

Performance on the practice task

A 2 x 2 mixed ANOVA was conducted to investigate the performance on the practice task of the two participant groups. The Levene's test was significant (easy task: $p < .001$; difficult task: $p = .051$), for that reason the data was log transformed (\ln) and reanalysed. The Levene's test remained significant (easy task: $p < .001$; difficult task: $p = .003$), therefore the data was analysed with the Mann-Whitney test. For both, the easy as well as the difficult morph task, a significant difference between the two groups was observed (easy task: $U = 219.00$, $p < .001$, $r = -.46$; difficult task: $U = 270.00$, $p = .011$, $r = -.33$). Control participants performed significantly better ($\underline{M} = 8.86$, $\underline{SD} = .23$, $Mdn = 18$) than the mLD group ($\underline{M} = 7.53$, $\underline{SD} = .23$, $Mdn = 16$). To examine further whether there were significant differences in performance for each participant group on the two different levels of task difficulty, the data was split for

group and two Wilcoxon signed-ranks tests were carried out. Both groups performed significantly better on the easy ($M = 8.99$, $SD = .15$, $Mdn = 9$) than on the difficult practice task ($M = 7.40$, $SD = .23$, $Mdn = 8$) (mLDs: easy task: $Mdn = 9$, difficult task: $Mdn = 7.5$; controls: easy task: $Mdn = 10$, difficult task: $Mdn = 8$) (mLDs: $T = 29.00$, $p < .001$, $r = -.49$; controls: $T = 27.00$, $p < .001$, $r = -.46$) (Table 10.3 displays means and standard deviations). These results replicate those obtained during Phase 1 of this study.

Table 10.3. Mean number of correct responses (hits) (+SD) during the practice tasks (easy and difficult) by mLD and control participants.

Correct responses during practice task				
	Easy task		Difficult task	
	M	SD	M	SD
mLDs	8.37	1.54	6.70	2.07
Controls	9.62	.62	8.10	1.45

Duration of the three construction phases

The duration (in minutes) of the three distinct composite construction phases were calculated for each participant to examine whether there were any group differences.

Means and standard deviations are displayed in Table 10.4.

Table 10.4. Mean durations of time (in minutes) (+SD) spent in the three different composite construction phases: selection and breeding (Phase 1), holistic tool (Phase 2) and shape tool (Phase 3), for both participant groups during the two description modes.

	Description Mode											
	Photo						Memory					
	Phase 1		Phase 2		Phase 3		Phase 1		Phase 2		Phase 3	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
mLD	20.57	5.15	8.28	2.46	4.56	3.53	21.4	7.5	8.3	4.17	4.5	3.13
Control	27.8	8.36	8.56	2.50	13.41	9.28	28.52	11.12	9.18	3.17	11.56	8.55

A between-subjects design MANOVA was carried out. The three dependent variables were the duration of the three construction phases. The two between-subject factors were group (mLD vs. control) and description mode (photo vs. memory). The Levene's test was significant for the duration of time participants spent in the shaping tool ($p = .005$). The data were log transformed (\ln) and reanalysed. The Levene's test remained significant ($p = .014$). Consequently, the data were collapsed across description mode and analysed with a Mann-Whitney test. It was found that participants with mLD spent significantly less time during the selection and breeding process and with the shaping tool than control participants (selection and breeding: $U = 257.50$, $p = .004$, $r = -.37$; shape tool: $U = 122.00$, $p < .001$, $r = -.63$) (mLDs: selection and breeding: $M = 21.1$, $SD = 6.8$, $Mdn = 21.35$; shape tool: $M = 4.31$, $SD = 3.33$, $Mdn = 3$) (controls: selection and breeding: $M = 28.00$, $SD = 9.52$, $Mdn = 25.48$; shape tool: $M = 12.49$, $SD = 9.5$, $Mdn = 9$). No significant group difference was revealed in the duration of time participants spent with the holistic tool ($U = 122.00$, $p = .22$, $r = -.16$) (mLDs: $M = 8.16$, $SD = 3.33$, $Mdn = 7.59$; controls: $M = 9.7$, $SD = 3.1$, $Mdn = 9$).

To investigate further what influence description mode had on the length of time participants spent in the different construction phases, the data were split for group and analysed with two separate Mann-Whitney tests. For both participant groups (control and mLD), there were no significant differences between the two description modes and the duration of time they spent in each construction phase (mLDs: photo: selection and breeding: $Mdn = 21.15$, holistic tool: $Mdn = 8.13$, shape tool: $Mdn = 3.14$; memory: selection and breeding: $Mdn = 21.5$, holistic tool: $Mdn = 6.43$, shape tool: $Mdn = 2.53$) (controls: photo: selection and breeding: $Mdn = 23.21$, holistic tool: $Mdn = 9$, shape tool: $Mdn = 10.59$; memory: selection and breeding:

$Mdn = 25.48$, holistic tool: $Mdn = 9$, shape tool: $Mdn = 7.34$) (mLDs: selection and breeding: $U = 111.00$, $p = .95$, $r < .001$; holistic tool: $U = 96.50$, $p = .507$, $r = -.09$; shape tool: $U = 95.00$, $p = .468$, $r = -.09$) (controls: selection and breeding: $U = 99.00$, $p = .575$, $r = -.07$; holistic tool: $U = 108.00$, $p = .852$, $r = -.02$; shape tool: $U = 91.00$, $p = .372$, $r = -.11$) (see Table 10.4 for means).

Acceptance of changes after the application of the holistic tool

After the use of the holistic tool, participants were asked whether they would like to keep or reject the changes made to the composite image. The number of mLD and control participants who accepted the applied changes was counted. To examine whether there was a significant association between group and whether participants accepted the changes the data were split for description mode and two Chi-square tests were performed. Both tables had cell frequencies less than 5, therefore the Fisher's Exact Test was used (as advised in Hinton, Brownlow, McMurray & Cozens, 2004, pp. 285). For the photo condition, the test approached significance ($p = .050$, one tailed Fisher's Exact Test). For the memory condition the test did not approach significance ($p = .299$, one tailed Fisher's Exact Test). Thus, during the photo condition, there was a strong association between group and whether or not participants accepted the changes made with the holistic tool. Seventy-three percent of the mLD participants accepted the changes, whereas all (100%) of the control participants accepted them. No significant association was obtained for the memory condition. Eighty percent of the mLD participants and 93% of the control participants accepted the holistic tool changes.

Amount of features changed with shape tool

The total number of facial features requested to be changed with the shape tool was calculated for each participant. The maximum number of facial features, which could have been changed, was 10: cheeks, ears, eyebrows, eyes, face width, forehead, jaw/chin/jowls, mouth/lips, nose/nostrils/temp and philtrum.

A 2 (group: mLDs vs. controls) x 2 (description mode: photo vs. memory) between-subjects design ANOVA was carried out to analyse the data. One significant main effect was identified for group $F(1, 56) = 37.45, p < .001, \eta^2 = .40$. Participants with mLD requested to change significantly fewer facial features ($M = 1.93, SD = 2.03$) than members of the control group ($M = 4.97, SD = 1.87$). The main effect for description mode $F(1, 56) = 3.80, p = .056, \eta^2 = .064$ and the interaction effect $F(1, 56) = .113, p = .738, \eta^2 = .002$ did not reach significance.

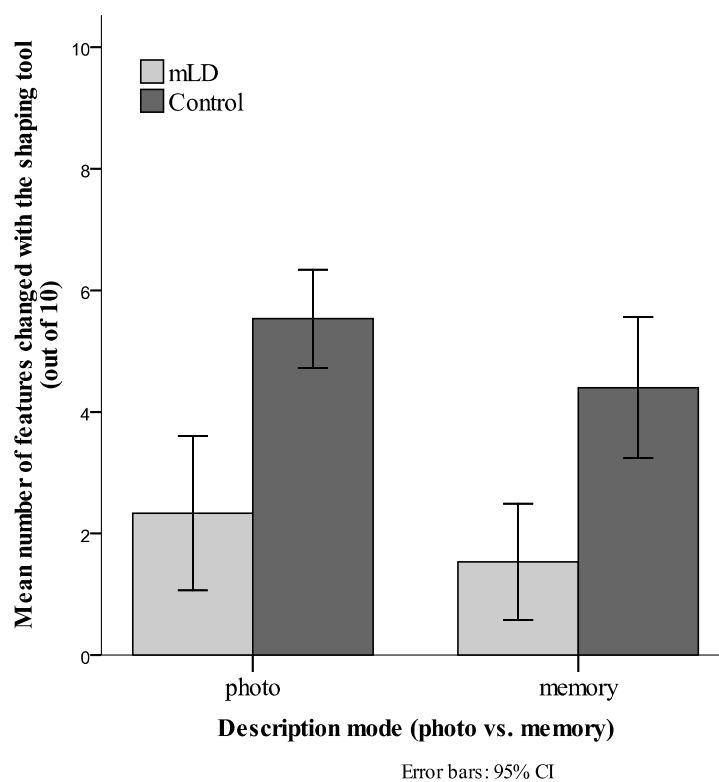


Figure 10.5. Mean number of facial features (out of 10) requested to be changed (+SE) for both participant groups during the two different description modes.

Relationship between WASI scores and performance

Correlations between mLD participants' WASI scores (full, verbal and performance WASI score) and their performance during the practice task and the composite construction phase were calculated to examine whether the intellectual functioning of mLD participants was related to their performance. Significant positive correlations were obtained between mLD participants' amount of correct responses during the easy and difficult morph task and their full- and performance-WASI scores (easy & full WASI score: $r = .544, p = .002$; easy & performance WASI score: $r = .528, p = .003$; difficult & full WASI score: $r = .435, p = .016$; difficult & performance WASI score: $r = .452, p = .012$). No significant associations were revealed between WASI scores and the time participants spent in the different EvoFIT construction phases (selection and breeding, holistic tool and shape tool) (all $ps > .05$) or the number of features requested to be changed with the shape tool ($p > .05$).

Summary of main findings

During the construction of facial composites with EvoFIT participants with mLD differed from participants without LD in several ways. First, they spent significantly less time with the selection and breeding process and the shape tool while creating the composite images. Moreover, mLD participants requested to change significantly fewer facial features with the shape tool than participants from the control group. No significant associations were revealed for the amount of accepted changes with the holistic tool and whether participants had LD or not. Finally, significant associations were revealed between mLD participants' full- and performance- WASI scores and the number of correct responses during the practice task. mLD participants with

higher WASI scores performed significantly better during the practice task than mLD participants with lower scores.

10.4 Part 3: Composite evaluation

During the composite evaluation the quality of the resulting composites was assessed with a matching task and a likeness rating. Participants not previously involved during the composite construction engaged in the evaluation part.

10.4.1 Matching task

Method

Design: A 2 x 2 mixed design was used, including one within-subject factor (group: mLD vs. controls) and one between-subject factor (description mode: photo vs. memory).

Participants: Forty-six participants (21 – 61 yrs; $M = 39.05$ yrs; $SD = 11.51$; 17 males and 29 females) took part. They were all members of the student and staff body from the University of Abertay Dundee.

Materials: The matching task consisted of the 60 EvoFIT composites, created during the composite construction and a 10 person line-up consisting of the five target faces and five distractor faces. The same target and distractor faces were used during the matching task in Study 3. Figure 10.6 shows an example of the matching task.

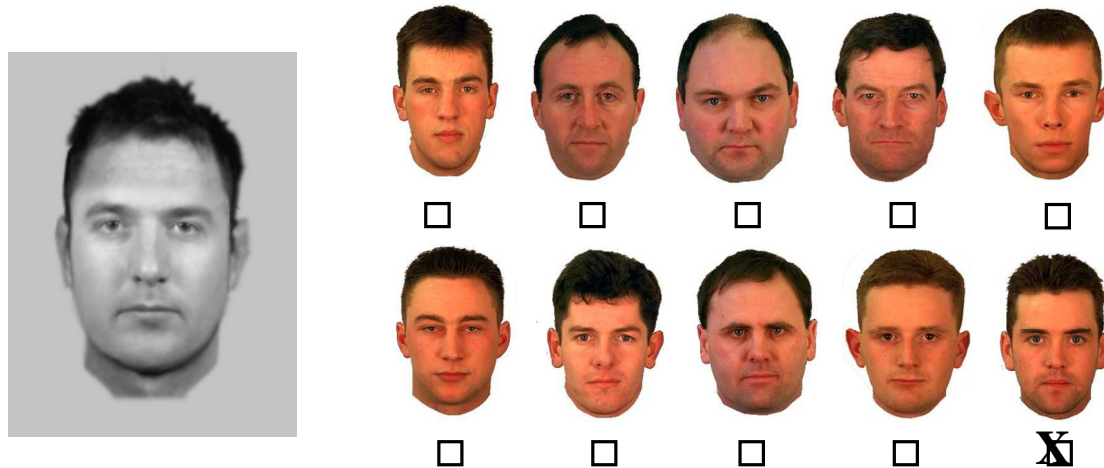


Figure 10.6. *Matching task: In this example, the participant has decided that the tenth face in the line-up (marked with an X) is most similar to the composite on the left.*

Procedure: The composites were presented to participants in booklets. Each booklet contained all 30 composites obtained during one of the two description modes (either photo or memory). Each composite was displayed on its own page and accompanied by a 10 person line-up on the next page. The order in which the composites were presented in the booklets was randomised as was the order in which the distractors and targets were presented in the line-ups. During the matching task, participants were asked to indicate which of the faces in the line-up best matched the accompanying composite. The dependent variable was the number of correct matches made on the basis of composites created by mLD participants and non-LD participants.

Results

For each participant the number of correct matches he/she made on the basis of composites created by mLD and control participants was calculated and the average scores are presented in Figure 10.7.

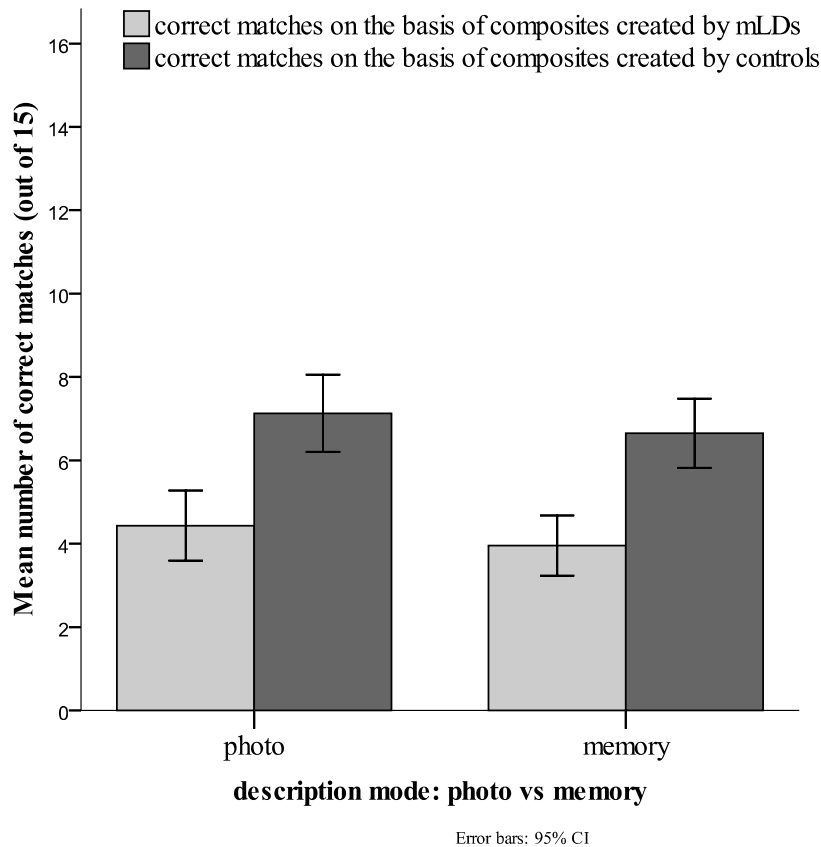


Figure 10.7. Mean number of correct matches (+SE) for composites created by mLD and control participants with EvoFIT during the two description modes.

A 2 x 2 mixed ANOVA was conducted to analyse the data. The within-subject factor was group (mLD composites vs. control composites) and the between-subject factor was description mode (photo vs. memory). The analysis revealed a significant main effect for group $F(1, 44) = 43.04, p < .001, \eta^2 = .49$. The main effect for description mode $F(1, 44) = 1.49, p = .229, \eta^2 = 0.33$ and the interaction effect $F(1, 44) = .00, p = 1.00, \eta^2 = .00$ were non-significant. Thus, participants made significantly more correct matches when the composites were created by members of the control group ($M = 6.89, SD = 2.02$), than when they were created by participants with mLD ($M = 4.1957, SD = 1.81$). There was no significant difference in the overall amount of correct matches for composites created in the photo condition ($M = 5.78, SD = .28$) and those constructed during the memory condition ($M = 5.30, SD = .28$).

To further investigate whether participants made more correct matches on the basis of facial composites constructed by mLD participants than by chance alone, the critical region for the whole group of participants was calculated. The binomial distribution had a mean of $pn = (1/10)(46 \times 30) = 138$ and a standard deviation of $\sqrt{npq} = \sqrt{1380(1/10)(9/10)} = 11.14$. To be significantly different from chance, the group score must be above (or below) the mean by at least $1.96(11.14) = 21.83$. Thus, with a mean of 138, the group would need to score above 159.83 ($138 + 21.83$) or below 116.17 ($138 - 21.83$) to be significantly different from chance performance. The group had a total score of 193 correct matches for composites generated by mLD participants, which is significantly above chance level. Thus, although participants had more difficulties accurately identifying the target person out of a 10-person lineup when the composite was created by an individual with mLD, they were nevertheless able to make significantly more correct matches than would have been expected by chance.

10.4.2 Likeness rating task

Method

Design: A 2 x 2 mixed design was used. The within-subject factor was group (mLD vs. controls) and the between-subject factor description mode (photo vs. memory).

Participants: Forty-six participants not involved in the previous composite construction phase and the matching task engaged in the likeness rating task. They were all students or staff from the University of Abertay Dundee (24 – 63 yrs; $M = 39.80$ yrs; $SD = 9.82$; 17 males and 29 females).

Procedure: The likeness rating task consisted of booklets including the 60 composites, created during the composite construction, alongside with the corresponding target faces. Participants were required to indicate how accurate each composite was by rating how well it resembled the target face on a Likert scale ranging from no similar likeness at all (1) to very similar likeness (10). See Figure 10.8 for an example of the likeness rating task.

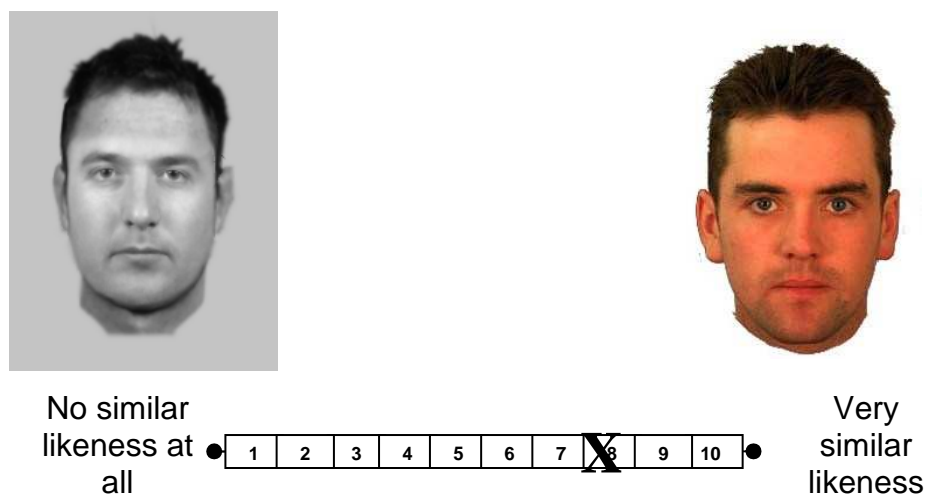


Figure 10.8. *Likeness rating task: In the example below, the participant has decided that the EvoFIT (on the left) is a quite similar likeness of the target (on the right); the participant has provided the EvoFIT with a score of 8 (marked with an X).*

Results

The average likeness rating scores for composites created by members of the mLD and control group were calculated and are displayed in Figure 10.9. A 2 x 2 mixed design ANOVA was conducted and obtained a significant main effect for group $F(1, 44) = 257.58, p < .001, \eta^2 = .85$. Thus, participants rated the composites created by the control group as significantly better likenesses ($M = 5.08, SD = 1.10$) than those constructed by the mLD group ($M = 3.55, SD = .92$). The main effect for description mode did not reach significance $F(1, 44) = 2.28, p = .138, \eta^2 = .05$ (photo: $M = 4.53, SD = .20$; memory: $M = 4.11, SD = .20$). However, a significant interaction between

group and description mode $F(1, 44) = 7.95, p = .007, \eta^2 = .15$ was obtained. Pairwise comparisons using the Bonferroni correction revealed that the group*description mode interaction derived from a significant difference in likeness rating scores between the photo ($M = 3.90, SD = .18$) and memory condition ($M = 3.21, SD = .18$) for composites created by the mLD group ($p = .010$). No such difference in likeness rating scores between the two description modes was present for composites created by the control group ($p = .646$) (photo: $M = 5.15, SD = .23$; memory: $5.00, SD = .23$) (see Figure 10.9).

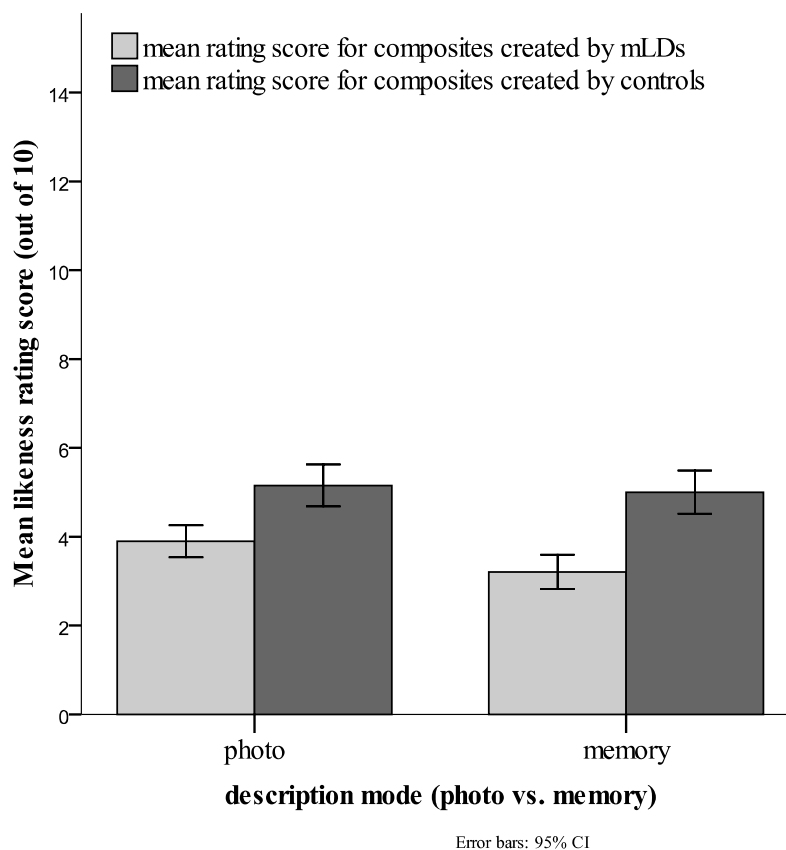


Figure 10.9. Mean likeness rating scores (+SE) for composites created by mLD and control participants during the two description modes.

Factors influencing composite quality

To investigate whether WASI scores, performance on the practice task, duration of composite construction and the number of features requested to be changed with the

shape tool had an significant impact on the quality of the resulting composites, the data obtained during the matching task and the likeness rating task were scored in an additional way, which is outlined in more detail in Chapter 8, pp. 197-198. As a result, each participant who has taken part during the composite construction part received for his/her composite image a matching score and a likeness rating score. These scores were correlated with the above mentioned variables of interest.

Matching task: The statistical analysis did not reveal any significant correlations between mLD participants' WASI scores (full-, verbal-, and performance-WASI score) and the quality of the created EvoFIT composites (all $ps > .05$). The performance on the practice tasks (easy and difficult morph task) was also not significantly correlated with the quality of the resultant composites (all $ps > .05$), neither for mLD participants nor for members of the control group. The same was true for the durations of the different EvoFIT construction phases (selection and breeding, holistic tool and shape tool) (all $ps > .05$). Finally, no significant association was obtained between the number of facial features requested to be changed with the shape tool and composites' quality ($p > .05$) (see Table 10.5).

Liikeness rating task: The likeness rating scores were correlated with the same variables outlined above and again all were not significant (all $ps > .05$).

Table 10.5. *Correlations obtained between participants WASI scores, performance on practice task, durations of composite construction phases, number of changed features with shape tool and the composite quality (Spearman's rho correlation coefficients and p values are provided).*

	mLD group				Control group			
	Matching task score		Likeness rating score		Matching task score		Likeness rating score	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>R</i>	<i>p</i>	<i>r</i>	<i>p</i>
Full WASI score	.048	.804	.146	.441	n.a.	n.a.	n.a.	n.a.
Verbal WASI score	-.095	.624	-.179	.345	n.a.	n.a.	n.a.	n.a.
Performance WASI score	.102	.597	.248	.187	n.a.	n.a.	n.a.	n.a.
Practice task score	-.115	.554	.176	.353	.136	.399	.240	.210
Selection and breeding phase	.116	.548	-.184	.332	.043	.821	.214	.257
Holistic tool phase	.362	.054	.185	.328	.198	.293	.325	.079
Shape tool phase	.305	.108	-.045	.813	.165	.385	.271	.148
Number of features changed	.064	.740	-.085	.655	.250	.183	.233	.216

When the matching task and likeness rating data of mLD and control participants were combined and correlated again with the above mentioned variables a different pattern emerged (see Table 10.6 for correlations). Strong correlations were obtained between composite quality and participants' performance on the practice task, durations of time participants spent in the different composite construction phases and the number of features participants requested to have changed with the shape tool. Nearly all of these correlations reached significance. The only two correlations which were only marginally significant, were between composite quality as assessed via the matching task and practice task performance ($p = 1.84$, $r = .177$) and duration of the selection and breeding phase ($p = .118$, $r = .206$).

Table 10.6. *Correlations obtained between participants performance on practice task, durations of composite construction phases, number of changed features with shape tool and the composite quality (Spearman's rho correlation coefficients and p values are provided).*

	mLD and control group combined			
	Matching task score		Likeness rating score	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Practice task score	.177	.184	.368	.004
Selection and breeding phase	.206	.118	.264	.042
Holistic tool phase	.294	.024	.283	.029
Shape tool phase	.351	.006	.399	.002
Number of features changed	.360	.005	.352	.006

Summary of main findings

In summary, the results of the matching task have revealed that participants had significantly more difficulties identifying the target out of a 10 person line-up on the basis of composites created by the mLD group compared to those created by control participants. The results obtained during the likeness rating task confirmed this finding. Thus, EvoFIT composites created by mLD participants were significantly poorer than those created by members of the control group. However, importantly, the findings of the matching task also showed that people with mLD were able to construct accurate composites with EvoFIT which could be used to identify a target face at better than chance performance by an independent sample of participants. It should be noted here that this was not the case with the E-FIT system tested in Study 3. The findings of this study further revealed that WASI scores of mLD participants were not significantly associated with the resulting EvoFIT composites' quality. However, participants' performance during the practice task, duration of time participants engaged with the construction of the composite and the amount of

features they changed with the shape tool were significantly correlated with the quality of the resulting composites, when collapsed across groups.

10.5 EvoFIT vs. E-FIT

Given that the same methodology, procedure, target faces and distractors were used in the E-FIT study outlined in Chapter 7, it is possible to combine the data from both experiments to investigate further which facial composite system is more suitable for people with mLD, E-FIT or EvoFIT. Therefore, the data sets of both studies were collapsed into one single data set, which allowed comparing the data obtained during the two evaluation tasks for each composite system (see Tables 10.7 for means and standard deviations during both description modes).

Table 10.7. Mean number of correct matches and likeness rating scores (+SD) obtained on the basis of E-FIT and EvoFIT composites created by either mLD or control participants during the two different description modes.

		E-FIT				EvoFIT			
		Matching task		Likeness rating		Matching task		Likeness rating	
		<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
mLDs	photo	3.78	1.28	2.65	1.22	4.43	1.95	3.90	.84
	memory	2.26	1.01	2.64	1.03	3.96	1.66	3.21	.89
Controls	photo	8.83	1.85	5.40	1.45	7.13	2.13	5.15	1.09
	memory	7.35	2.66	5.15	1.29	6.65	1.92	5.00	1.13

10.5.1 Matching task

A 2 x 2 x 2 mixed design ANOVA was conducted. The two between-subjects factors were program (E-FIT vs. EvoFIT) and description mode (photo vs. memory). The within-subject factor was group (mLD vs. control). A significant main effect for group was observed $F(1, 88) = 193.95, p < .001, \eta^2 = .69$, with participants making

significantly more correct matches when the composites were created by the control group ($M = 7.49$, $SD = 2.28$) than by participants with mLD ($M = 3.61$, $SD = 1.70$). In addition, a significant main effect for description mode was revealed $F(1, 88) = 13.16$, $p < .001$, $\eta^2 = .13$, demonstrating that participants made considerably more correct matches on the basis of composites created during the photo condition ($M = 6.04$, $SD = .19$) than during the memory condition ($M = 5.05$, $SD = .19$). The main effect for program was not significant $F(1, 88) = .002$, $p = .968$, $\eta^2 = .13$, there was no significant difference in the overall number of correct matches for composites created with E-FIT ($M = 5.55$, $SD = .19$) or EvoFIT ($M = 5.54$, $SD = .19$).

Intriguingly, the analysis revealed a significant interaction effect between group and program $F(1, 88) = 18.08$, $p < .001$, $\eta^2 = .17$. To explore the significant interaction effect in more depth, pair wise comparisons were carried out using the Bonferroni correction. The post hoc analysis revealed that the significant interaction arose because participants had significantly more correct matches when the composites created by the mLD group were constructed with EvoFIT ($M = 4.20$, $SD = .22$) than with E-FIT ($M = 3.02$, $SD = .22$) ($p < .001$). However, the opposite was true for composites created by the control group, participants had significantly more correct matches when those were created with E-FIT ($M = 8.09$, $SD = .32$) than with EvoFIT ($M = 6.89$, $SD = .32$) ($p = .010$) (Figure 10.10 displays the interaction).

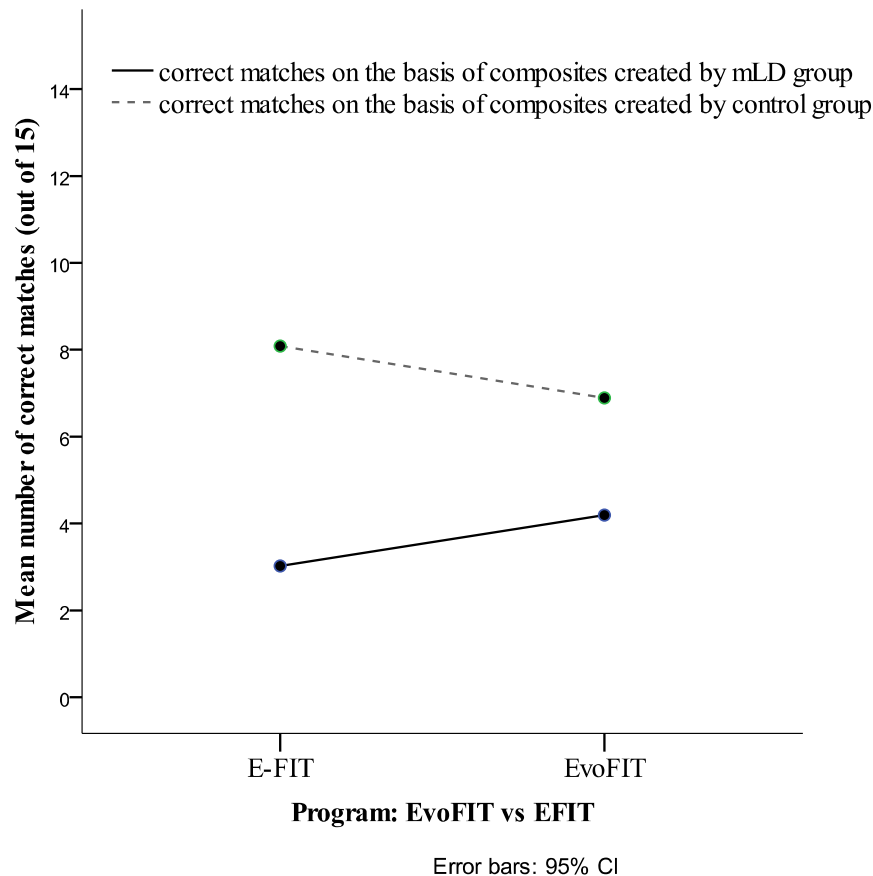


Figure 10.10. Mean number of correct matches on the basis of composites created by the mLD group and the control group with E-FIT and EvoFIT (the data was collapsed across description mode).

The remaining two-way interactions between group and description mode $F(1, 88) = .002, p = .969, \eta^2 = .00$ and program and description mode $F(1, 88) = 3.51, p = .064, \eta^2 = .13$ were non significant, as was the three-way interaction between group, program and description mode $F(1, 88) = .002, p = .969, \eta^2 = .00$.

10.5.2 Likeness rating task

A 2 x 2 x 2 mixed design ANOVA was carried out. As during the analysis of the matching task data, the two between-subject factors were program (E-FIT vs. EvoFIT) and description mode (photo vs. memory). The within-subject factor was group (mLD vs. control). A significant main effect for group was obtained $F(1, 88) =$

587.83, $p < .001$, $\eta^2 = .87$, indicating that participants rated composites created by the control group ($M = 5.18$, $SD = .13$) as significantly better likenesses than those created by the mLD group ($M = 3.10$, $SD = .11$). The main effects for program $F(1, 88) = 2.63$, $p = .108$, $\eta^2 = .03$ and description mode $F(1, 88) = 1.53$, $p = .219$, $\eta^2 = .02$ did not reach significance. A significant interaction effect between group and program $F(1, 88) = 41.52$, $p < .001$, $\eta^2 = .32$ and a significant three-way interaction between group, description mode and program $F(1, 88) = 5.14$, $p = .026$, $\eta^2 = .05$ were obtained. The two-way interactions between group and description mode $F(1, 88) = .74$, $p = .392$, $\eta^2 = .008$ and description mode and program $F(1, 88) = .45$, $p = .506$, $\eta^2 = .00$ did not reach significance.

To explore the significant interaction effects further, pair wise comparisons were performed with the Bonferroni correction. The significant two-way interaction between group and program arose because participants awarded EvoFIT composites ($M = 3.55$, $SD = .15$) with significantly higher likeness scores than E-FIT composites ($M = 2.64$, $SD = .15$), when those were created by mLD participants ($p < .001$). However, for composites created by the control group, no significant difference in the likeness rating scores between composites created with E-FIT ($M = 5.27$, $SD = .18$) and EvoFIT ($M = 5.08$, $SD = .18$) were obtained ($p = .45$).

The significant three-way interaction derived from a significant difference between the two description modes in the likeness rating scores for composites created with EvoFIT by mLD participants ($p = .023$). Participants rated EvoFIT composites created by the mLD group as significantly better likenesses when those were constructed during the photo condition ($M = 3.40$, $SD = .21$) than during the memory condition ($M = 3.20$, $SD = .21$) (Figure 10.11 displays the three-way interaction).

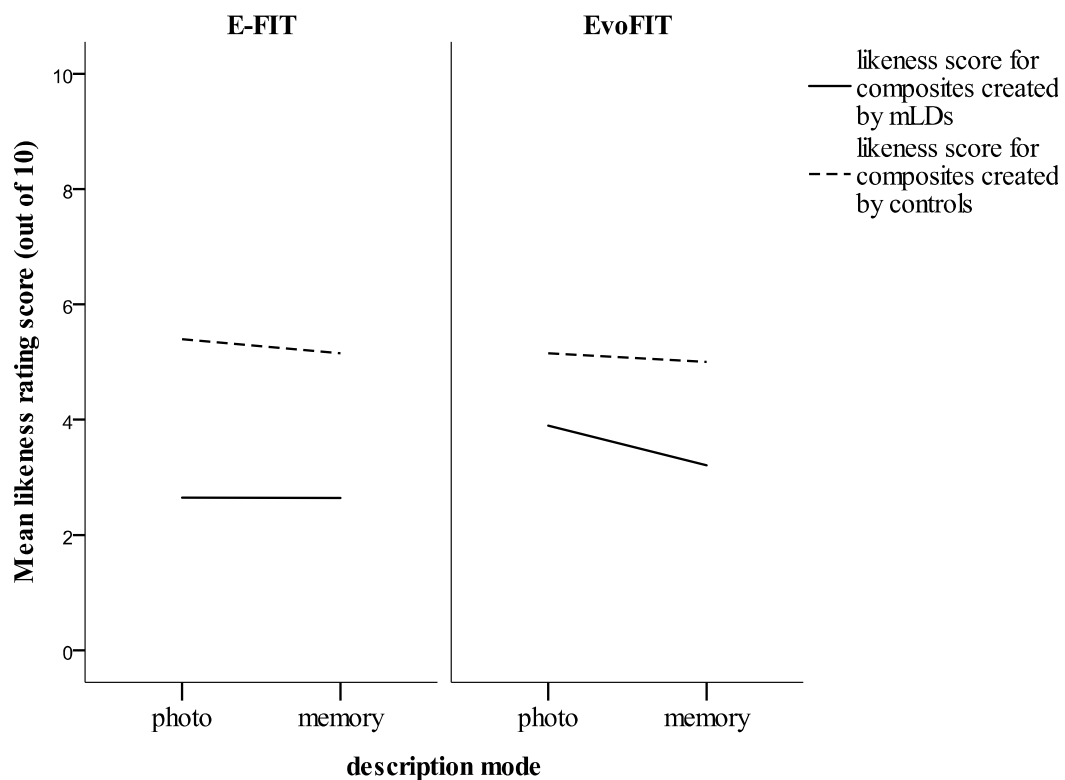


Figure 10.11. Mean likeness rating scores for composites created by mLD and control participants with E-FIT and EvoFIT during the two description modes. The three-way interaction arose, because description mode had only a significant impact on mean likeness rating scores for composites created by mLD participants with EvoFIT.

10.5.3 Summary of main findings

The statistical comparison of the results obtained during the matching and the likeness rating tasks revealed that regardless of facial composite system utilised, composites constructed by participants with mLD were significantly poorer than those constructed by participants without LD.

Furthermore, in both evaluation tasks, people with mLD created significantly better composites with EvoFIT than with E-FIT. A different picture emerged for composites created by control participants. Specifically, the comparison of the matching task data showed that control participants had significantly more correct matches when the composites were created with E-FIT than with EvoFIT. No

significant difference in quality between E-FIT and EvoFIT composites was revealed during the comparison of the likeness rating tasks for control participants. However trends were revealed which point into the same direction.

With regard to description mode, the comparative analysis of the matching task data demonstrated that regardless of composite system, composites were significantly better when created with the target photograph in view than from memory alone. The comparison of the likeness rating data obtained only a significant difference between the two description modes for composites created by mLD participants with EvoFIT.

10.6 Discussion

The findings of previous studies during this PhD-thesis have demonstrated that individuals with mLD consistently perform poorer than people without LD on tasks involving face recognition and description. Therefore, the expectation of the current study was not to demonstrate that with the use of novel holistic facial composite systems, such as EvoFIT, people with mLD can perform better than control participants, but simply to investigate whether they can be regarded as a potential tool to improve the performance of individuals with mLD, maybe to a level equivalent to that of people without LD.

The findings of the morph task showed that at least the majority of people with mLD had the ability to differentiate between more and less similar looking faces and to understand abstract terminology, such as the concept of similarity, which might be used by the operator during the composite construction with EvoFIT. This finding opened up the possibility of conducting the EvoFIT study to investigate whether it might be a possible alternative to the older E-FIT system for witnesses with mLD.

During the composite construction with EvoFIT, participants with mLD differed from participants without LD in a number of ways. First, as in the E-FIT study, participants with mLD spent less time with the composite construction than members of the control group. This finding suggests that participants without LD took significantly longer to make their similarity judgments during the selection and breeding phase and to edit the final composite image with the shape tool than members of the mLD group. The latter was also confirmed by the finding that mLD participants requested to change significantly fewer facial features with the shape tool than members of the control group, which is again in agreement with the results obtained during the E-FIT study.

No significant group differences were obtained in the length of time participants used the holistic tool set and whether they subsequently accepted the applied holistic changes or not. On the basis of feedback provided by participants, it could be argued that the rationale behind the holistic tool was difficult to comprehend, particularly by individuals with mLD, and that they did not experience it as a valuable tool to enhance the composite likeness. This was confirmed by the finding that participants spent significantly more time in the selection and breeding phase and in the shape tool phase than in the holistic tool phase. Further evidence for the suggestion that the holistic tool was particularly difficult to use for people with mLD derives from the finding that only 76.7% of mLD participants accepted the applied holistic changes, whereas 96.7% of control participants did. The subjective opinion of participants during the present study is however in disagreement with research findings which have revealed a clear benefit for the holistic tool (Frowd et al., 2007) when used with participants without LD. Frowd et al. (2007) demonstrated that the resultant EvoFIT composites were significantly better named when they were

previously modified with the holistic tool. The discrepancy of these findings implies that although the holistic tool may be effective in increasing the likeness of composite images, it does not seem to be particularly user friendly, specifically not for individuals with mLD.

All the observed group differences during the EvoFIT composite construction process may have contributed to the differing quality of the resultant composites. Support for this assumption derives from the correlational analysis, that showed that composite quality is strongly related to participants' performance during the practice task, the duration of time they engage with the composite construction and the amount of features modified with the shape tool. The results of the evaluation tasks showed that the quality of EvoFIT composites was significantly poorer for those created by the mLD group than for those created by the control group. Thus, it was not possible to increase the composite construction performance of people with mLD to an equivalent level of those without LD by using a holistic facial composite system. It was however possible to enhance their performance significantly compared to that revealed with the earlier E-FIT system during Study 3. Potential reasons for the enhanced composite quality might be that the program is less reliant on a verbal description of the target face. Instead, it relies more on face recognition, which theoretically should be a considerable benefit for people with mLD. Research has shown that they have limited verbal abilities (Agnew & Powell, 2004; Emerson, 2001; Milne, Clare & Bull, 1999) but reasonable face recognition abilities (Dobson and Rust, 1994). Furthermore, the EvoFIT system requires every witness to go through the same fixed composite construction procedure. Each of the three composite construction phases (selection and breeding, holistic tool and shape tool phase) needs to be completed before the EvoFIT construction process is finished. This means it is

not entirely up to the witness to decide when the composite construction process is finished, unlike with the E-FIT system. This might be advantageous for individuals with mLD. For example, Study 3 revealed that they express satisfaction with the resultant composite image at an earlier stage than do control participants and that their resultant composites were judged as significantly poorer likenesses. Furthermore, the findings of this study demonstrated that longer engagement with the composite construction is associated with better composite quality. Therefore, giving witnesses with mLD less freedom of choice and more time to work on the composite image might result ultimately in superior likenesses.

The results also suggest that the poorer E-FIT composites generated by mLD individuals in Study 3, were likely due to problems with both memory and language. If the poor quality of the facial composites created with E-FIT was merely due to problems with language, one would have expected mLD participants' performance with a program relying less on language to be similar to that of individuals without LD. This was not the case, implying that poor composite quality is probably a result of both memory and language deficiencies in individuals with mLD. However, firm conclusions should be drawn with caution. Problems with communication and language might not only manifest themselves in terms of difficulties to verbally describe faces or events, but also in terms of problems with comprehending task instructions and effectively interacting with others. Although EvoFIT relies less on language abilities than E-FIT, in that it does not require witnesses to verbally describe the perpetrator's face, there is still need for communication between the operator and the witness during the composite construction.

Intriguingly, EvoFIT did not appear to be superior for participants without LD. This finding is in agreement with earlier research, which has employed similar time

delays (2-4 hour delays) between target face presentation and facial composite construction (Frowd et al., 2005a; 2005b). However, it contradicts the findings of more recent studies which have employed considerably longer time delays (two-day delays) (Frowd et al. 2007; 2010). These later studies have demonstrated that EvoFIT can produce composite images which are superior to those generated with previous computer-based systems. Frowd et al. (2007) argued that the different evaluation tasks employed during earlier and more recent studies might be responsible for the discrepancies in findings. They claim that feature-based evaluation tasks, such as matching and sorting tasks are not the most appropriate way to assess the performance of a holistic composite system. Instead, evaluation tasks based on holistic face processing such as naming tasks should be regarded as more appropriate to assess the utility of EvoFIT. According to Frowd et al. (2007), these tasks should also reveal more favourable results for the EvoFIT system. However, there is no clear evidence for Frowd et al.'s (2007) justification. Although the most recent studies (Frowd et al. 2007; 2010) have demonstrated that holistic systems can produce better-quality composites than computer-based systems, when a naming task served as evaluation tool; mixed results were obtained in earlier studies, which have employed both holistic as well as feature-based evaluation tasks (Frowd et al., 2005a; 2005b).

A possible alternative explanation for the contradicting findings obtained in previous and more recent studies may be that the exposure duration of the target face and the delay between target and subsequent composite construction determine whether people engage in a more holistic or featural face-processing approach. Depending on the specific face-processing approach, either holistic or computer-based systems will produce better likenesses. Evidence for this hypothesis comes from a study conducted by Hole (1994). During a series of experiments, Hole investigated

the face inversion effect (also known as the composite effect first demonstrated by Young et al. in 1987) with unfamiliar faces and revealed that slight alterations in the experimental procedure can considerably affect participants' face matching strategies. During the study, Hole (1994) presented participants with pairs of unfamiliar chimeric faces, which consisted of the top half of one face and the bottom half of a different one. Participants were required to decide as quickly as possible whether or not the top halves were identical. The stimuli were presented either inverted or in an upright orientation. Inversion is thought to disrupt holistic face processing thereby reducing the chimeric effect. During the first experiment, Hole (1994) presented the stimuli for two seconds, a procedure that did not reveal the expected face inversion effect. Thus, participants were not quicker at making judgements for chimeric faces which were presented inverted than for those presented in an upright orientation. On the basis of participants' feedback, Hole (1994) argued that people engaged in a feature-by-feature matching strategy and therefore performance was not hindered by the upright orientation. The follow-up experiment replicated the procedure of the first one, except that the exposure duration of the stimuli was considerably shorter, i.e. 80 milliseconds. This time the face inversion effect was revealed, participants were significantly quicker in matching inverted face halves than upright ones. Thus, the very short exposure duration during the second experiment seemed to have prevented participants from adopting a featural face-processing approach, which they have used successfully in experiment one, and encouraged them to engage in more holistic face-processing. The results of this study clearly demonstrate how slight alterations in the procedure can impact on peoples' face-processing strategy.

Additional support for the assumption that people adopt different face processing strategies (featural vs. holistic), depending on the specific task

requirements was provided by Farah, Drain and Tanaka (1995). They asked participants to study names associated with faces. The faces were either presented holistically (as a whole face) or in their individual parts (the head outline, nose, mouth, and eyes). During the test phase, participants were presented with the studied faces (all of them were presented in a holistic manner), which were either displayed inverted or upright and they were asked to name them as quickly as possible. Farah et al. (1995) found that the face inversion effect was only present in the performance of those participants who studied the holistic faces. Participants who studied the face parts did not display the inversion effect. Thus, Farah et al. (1995) successfully manipulated participants to engage in a featural-face encoding strategy, which prevented them from engaging in holistic face processing at the test phase.

With regard to the present study, it could be argued that the relatively long exposure of the target face (one minute) and the specific task demands encouraged participants to engage in a more featural-face encoding approach and this subsequently lead to better performance with a composite system congruent with this encoding strategy, i.e. E-FIT. However, the question arises why other studies have obtained results in favour for EvoFIT, although they have applied similar target exposure durations (Frowd et al. 2007; 2010). A potential reason for this inconsistency in findings might be the duration of the delay applied between the presentation of the target and the composite construction. It is possible that a longer delay weakens the featural representation of the face in memory and individuals are left with a more holistic, however, less detailed impression of the target face. This representation change in memory might account for the discrepancy in findings between studies which have used similar target exposure durations but different durations in delay. During the current study, participants inspected the target faces for

a relatively long duration (one minute), however, the delay (three hours) was rather short compared to other studies, which have employed two-day delays (Frowd et al. 2007; 2010). Therefore it can be argued that during the present study, participants constructed the composites with a more featural-face representation in mind, which was more consistent with the composite construction approach of the E-FIT system than with the holistic one utilised by EvoFIT.

Further support for the assumption that delay can modify memory representations derives from the Fuzzy-trace theory (FTT) (Reyna & Brainerd, 1995), which was originally used to account for the role memory plays in higher reasoning processes and later to explain false memory effects (Brainerd & Reyna, 2002). According to FTT, people acquire verbatim and gist memories about an experienced event and these memories are stored separately (Reyna & Titcomb, 1996). Verbatim memories refer to memories of the surface forms of experienced items, whereas gist memories refer to memories regarding meaning, relations and patterns. Whether people assess the verbatim or gist representations is influenced by delay and associated forgetting (Reyna & Titcomb, 1996). Forgetting is more rapid for verbatim than for gist representations. Consequently, people rely more on verbatim memories during immediate memory tests, however, they shift to gist representations after a delay (Reyna & Kirnan, 1994). Evidence for FTT stems from studies investigating autobiographical memory in children (Poole & White, 1993), memory for numbers (Brainerd & Gordon, 1994) and memory for sentences (Reyna & Kirnan, 1994). The findings of the current study suggest that similar principles might apply to memory for faces. The more featural-face processing approach would also explain why control participants wanted to change a considerable amount of facial features with the shape

tool, after having already successfully completed the selection and breeding and the holistic tool phases.

The question arises, why people with mLD have nevertheless benefited from the usage of a holistic facial composite system, although it is actually not in agreement with their employed face encoding strategy and subsequent memory representation of the target. It could be argued that participants with mLD profited from the EvoFIT system more, not because of its holistic nature, but because of it being less dependent on language than the E-FIT system. This advantage might have outweighed the discrepancy in processing approach and therefore mLD individuals produced better composites with EvoFIT than with E-FIT. Thus, depending on the specific abilities and disabilities of the participant group (in this case people with mLD and without LD) and the face-encoding strategy applied, either more featural or more holistic composite systems can be regarded as more effective.

Another interesting finding obtained during the present study was that unlike the results obtained during the E-FIT study, no significant difference in EvoFIT composite quality was obtained between the two description modes: photo vs. memory. It could be reasoned that the EvoFIT system does not benefit to the same degree from a detailed verbal description of the target face as the E-FIT system. Consequently, no considerable advantage was revealed when the target face was in view during the composite construction. However, trends point into the expected direction, with better composites created during the photo than the memory condition.

As observed in earlier studies described in this thesis, intellectual functioning of mLD participants, as assessed via the WASI, was not correlated with performance during the composite construction nor with the quality of the resulting composites. This casts further doubts on the effectiveness of the WASI to predict in some way the

performance of individuals with mLD in an eyewitness setting, such as during the construction of facial composites. However, other variables appeared to have more predictive value. For instance, people created significantly better rated composites when they performed well during the practice task. Furthermore, the quality of the composites was higher when participants engaged longer with their construction. Finally, the more features participants requested to change with the shape tool at the end of the construction process the more resembled the resultant composite the target face. These findings have important practical implications, such as that the practice task could serve as a screening tool to assess composite construction abilities, particularly when the police is dealing with witnesses and victims with LD and time constraints do not allow for a detailed assessment of the individuals specific abilities and disabilities.

In conclusion, the results of the present study suggest that EvoFIT can be regarded as a more suitable composite system for witnesses with mLD, as opposed to more featural systems, such as E-FIT. The underlying reasons for this finding might be that EvoFIT relies less on language than E-FIT. Moreover, it provides witnesses with less freedom of choice, which might be particularly beneficial for witnesses who are highly suggestible and prone to acquiesce, such as individuals with mLD. Furthermore, the findings provide novel theoretical ideas regarding how humans in general process and remember faces, such as that the duration of exposure and delay can influence whether people engage in more featural or holistic face processing strategies. It should be acknowledged though, that these are at present speculative.

Chapter 11

General discussion and concluding remarks

This final chapter provides an overview of the main findings obtained during the experimental studies described in Chapters 6 to 10 and their theoretical and practical contributions are discussed. Furthermore, methodological limitations of the conducted research are addressed and future research directions considered. Finally, the most relevant practical implications for forensic settings are highlighted.

11.1 Introduction

During this thesis, the ability of people with mLD to recognise, describe, and create facial composites of unfamiliar faces was investigated. These abilities can be of particular importance in eyewitness situations, such as during the construction of facial composites with a police operator. Due to their high prevalence rate and increased vulnerability to victimisation people with mLD might be more likely to get into such situations than other members of the wider community (Emerson, 2001; Nettelbeck & Wilson, 2002). It was argued therefore that it is important to conduct research in the area of face recognition and recall in people with mLD, which has been to date somewhat neglected in the forensic research literature. The first study, described in Chapter 6 in this thesis, is a survey study which assessed current usage of composite systems by UK police operators and their attitudes towards and experiences with LD witnesses. Overall, the aim of this survey study was to investigate whether the intended research objectives were indeed of practical relevance to the police. Thereafter, the experimental studies followed. The studies progressed from the investigation of basic face recognition and description abilities in

individuals with mLD to the exploration of more applied aspects such as facial composite construction skills. In addition, potential measures to facilitate composite construction with mLD witnesses were examined. The findings of the studies conducted within this thesis contribute to the existing base of knowledge regarding eyewitness performance amongst individuals with mLD (reviewed in Chapter 3) and to the understanding of human face processing in general (reviewed in Chapter 4). Moreover, the results add to the existing research literature regarding the evaluation of contemporary facial composite systems (reviewed in Chapter 5). Finally, the findings fill the gap of knowledge regarding face recognition and recall abilities in individuals with mLD. The main findings of these studies are summarised below, followed by discussions of both their theoretical and applied implications.

11.2 Summary and main findings

11.2.1 A survey of facial composite operators

The survey study set the context for the following experimental studies. Most importantly, it confirmed the need for an investigation of face recognition and recall abilities in individuals with LD. One-third of the police operators who completed the survey had previous experience in generating facial composites with witnesses with LD. Thus, the original assumption that people with LD may find themselves in the situation where they have to describe a perpetrator's face to the police is supported by evidence that this does indeed happen. Moreover, the survey revealed that two of the most frequently used facial composite systems in the UK appear to be E-FIT and EvoFIT. These findings support the decision to concentrate specifically on these two systems in this thesis. In addition, the study revealed that most police operators are not aware of any specific guidelines to which they could refer to when creating

composite images with LD witnesses. This finding supports the lack of research regarding face recognition and recall in individuals with LD, and implies that there may be no specific guidelines available regarding the effective generation of facial composites with LD witnesses. Taken together, the findings from the survey emphasised the need for future research in this area.

11.2.2 Face recognition and description abilities in people with mLD of unfamiliar faces

Study 2 investigated basic face recognition and description abilities in people with mLD, and can therefore be regarded as a central building block for the experimental studies that followed. The first part of the study established that people with mLD are able to accurately remember, and subsequently recognise, unfamiliar faces. However, their performance was considerably poorer than that of control participants. Furthermore, mLD participants also needed additional practice before being able to manage the task better than would have been expected by chance alone.

The second part of the study produced evidence that people with mLD have considerable difficulties in providing a detailed and accurate verbal description of an unfamiliar face. Their facial descriptions were significantly less detailed and accurate than those provided by control participants. An important finding is that, in line with the performance in the recognition task, participants with mLD showed variability in their performance on the description task. This finding was consistent with that observed in individuals without LD.

11.2.3 The efficiency of E-FIT with mLD witnesses

After having established that individuals with mLD are able to recognise and describe previously seen unfamiliar faces, Study 3 investigated these abilities in a more applied

setting; during the construction of facial composites with E-FIT. The findings of this study indicate that individuals with mLD generated significantly poorer composites than members of the control group. Their inferior performance may be explained by their sparse verbal, as well as non-verbal facial descriptions, fewer requested alterations to their composite image, and the ease by which they were satisfied with the resultant composite. Notwithstanding these findings, some of the facial composites created by people with mLD were of a high enough standard that an independent sample of participants was able to identify the target faces on the basis of them out of a 10 person line-up.

11.2.4 Do visual prompts facilitate verbal descriptions of unfamiliar faces in witnesses with mLD?

During Study 4, visual prompts were used to facilitate the description of faces by individuals with mLD. However, participants with mLD seemed to select the visual prompts in a rather random fashion compared to control participants. This was evidenced by only moderate agreement between the prompts selected by mLD individuals. Furthermore, the prompts selected by mLD participants were often in disagreement with those chosen by the control group. As a result, the use of these prompts during the following experimental studies was not adopted. Instead, it was regarded as more profitable to focus on facial composite systems that rely less on verbal abilities of the witness.

11.2.5 The suitability of EvoFIT for mild learning disabled witnesses

Study 5 investigated the performance of mLD individuals with EvoFIT, a novel holistic facial composite system. The main aim of this study was to assess whether people with mLD are able to use the EvoFIT system, and whether it may be a

potential alternative to the older featural-system, E-FIT. Our findings revealed that mLD individuals constructed poorer EvoFIT composites than control participants. The difference in composite quality may be due to several factors, such that mLD participants spent less time with the whole composite construction procedure and amended significantly fewer facial features during this process. However, most importantly, the study revealed that mLD individuals were able to create significantly better composites with EvoFIT than with E-FIT.

11.3 Discussion of main findings

A consistent finding across the experiments conducted during this PhD thesis was that participants with mLD performed poorer on tasks involving face recognition and recall than participants without LD. This finding was reliable across a wide range of stimuli (Mr Men characters, male and female target faces, and morphs) and tasks (forced choice recognition task, description tasks, including free recall and cued recall phases, the construction of facial composites with different composite construction techniques, such as E-FIT and EvoFIT), and tend to fit the layman's viewpoint that individuals with mLD are less reliable eyewitnesses (Stobs & Kebbell, 2003; Peled et al., 2004). However, although mLD individuals performed poorer than their non-LD peers, they were able to complete most of the tasks above the level expected by chance. Most importantly, the studies in this thesis reveal that mLD participants exhibit variability in their face recognition and recall performance dependent upon the measures introduced. Interestingly, they sometimes benefit from the same measures as individuals without LD benefit from. Conversely, on other occasions, they profited from measures which were not ideal for non-LD individuals. The observed variability in performance and the differences between the two participant groups provide

important insights into establishing why mLD individuals' performance was poorer and what measures could be put in place to improve their performance in such tasks.

11.3.1 The ability of people with mLD to recognise and verbally describe faces

Part 1 of Study 2 shows that people with mLD are significantly less accurate in recognising previously seen unfamiliar faces. A finding consistent with previous research regarding mLD individuals' face recognition abilities, demonstrating that their performance level overall is lower than that of control participants (McCartney, 1987). However, most importantly, the study reveals that mLD individuals' performance improves with increased practice. Whilst their performance is below chance on the first old/new face recognition task it improves to above chance level on practice task 2 and during the main recognition task. This positive effect of practice is in line with earlier research in the area of metacognition. Several studies show that individuals with LD can be trained to use rehearsal strategies, which can subsequently improve short-term memory performance (Belmont & Butterfield, 1971; Butterfield, Wambold & Belmont, 1973; Brown & Barclay, 1976). For instance, Belmont and Butterfield (1971) required participants with and without LD to learn letter lists and to accurately recall the serial letter positions at test. They argued that the superior performance at test of participants without LD was a result of their greater rehearsal during the learning phase, as evidenced by increases in pausing between letters when the information load increased. Such a gradual increase in pausing was not observed for LD participants. Conversely, their pausing decreased, indicating that they did not engage in similar amounts of rehearsal as their non-LD counterparts. However, when providing participants with explicit effective rehearsal instructions, both groups

showed a considerable increase in later memory performance. Nevertheless, in line with what is reported in this thesis, performance of LD participants did not reach the level of their non-LD peers. Similar beneficial effects of training were observed in studies regarding social skills in LD individuals (see McIntosh, Vaughn & Zaragoza, 1991 for a review) and the acquisition of everyday routines, such as the preparation of food and the detection of potential hazards in the work environment (Brooks, Rose, Attree & Elliot-Square, 2002). Apparently, even very simple forms of training such as the mere repetition of to-be-remembered material can enhance performance in people with and without LD, as demonstrated by Henry and Gudjonsson (2004). They asked children with and without LD to watch a short video-clip of a minor crime. After a short delay (three to four 4 minutes) children were interviewed about the event. Half of the children watched the video-clip twice prior to the interview session. It was revealed that viewing the video-clip for a second time dramatically increased the accuracy and quantity of recalled information in individuals with and without LD. Thus, the present findings from this thesis suggest that people with mLD do not suffer from a general defect in face processing, as evidenced by the finding that they were able to manage the face recognition tasks, after sufficient training, at a level above chance. However, they may engage in less efficient learning strategies which subsequently result in poor recognition performance.

The results, obtained during Part 1 of Study 2, have important practical implications for the police and other practitioners working with LD individuals. First, people with mLD may already have difficulties understanding simple instructions and tasks which normally developed individuals have no problems with. These may include picking a previously seen face out of a two person array. This might be particularly important in an eyewitness setting, when the witness is required to select

the suspect from an identification parade or a mug shot book. Secondly, providing LD individuals with additional instructions to support them in how to best complete the task at hand, or with the opportunity to practice the task (for example with a different set of stimuli or to see somebody else complete the task) may have a beneficial effect on their subsequent performance.

Apart from mLD individuals' face recognition performance, Study 2 also examined their ability to describe previously seen faces. The findings reveal that the verbal facial descriptions provided by mLD individuals are considerably less complete and less accurate than those provided by their non-LD peers. This is partially in line with research regarding eyewitness event recall, which repeatedly demonstrates that recall by LD individuals is overall less complete than that provided by people without LD (Agnew & Powell, 2004; Brown & Geiselman, 1990; Henry & Gudjonsson, 2004; Michel et al. 2000; Perlman et al., 1994). However, contrary to the present findings, most studies have found no significant group differences with regard to the accuracy of the obtained event information (Henry & Gudjonsson, 1999; Henry & Gudjonson, 2004; Michel et al., 2000; Perlman et al., 1994). Thus, it seems as if the task to verbally describe a face can be considered as more challenging than recalling the course of actions during an experienced event, particularly for those individuals with LD. The notion that people in general experience difficulties when describing faces per se and their individual features is not a new one (as reviewed in Chapter 4). Archival studies have shown that facial descriptions are often vague and rare (Van Koppen & Lochun, 1997).

Intriguingly, although people with mLD perform overall at a lower level on the face description task compared to members of the control group, they nevertheless show the same response pattern during the different description modes and recall

conditions. Thus, mLD individuals benefit from the same introduced measures as individuals without LD. Both groups mention considerably more facial information when interviewed with a cued-questioning approach and when the photo of the target face was in view, compared to the free recall condition, involving the description of the target face from memory alone. With regard to the accuracy of the obtained facial information, no significant differences are revealed between the different description modes and the recall conditions. However, although not significant, the observed trends point into the direction that both participant groups provide more accurate information during the free recall with the target in view. The information provided decreases in accuracy with the introduction of more specific questions and when the information was recalled from memory. This pattern of performance is in agreement with previous eyewitness event recall research in relation to open and closed questions (Dent, 1986; Henry & Gudjonsson, 1999; Michel et al. 2000; Perlman et al. 1994). It appears reasonable that people provide more information when they described the target with the photo in view, as this condition can be regarded as less cognitively demanding than describing the face from memory. Furthermore, it could be argued that people provide significantly more information during the cued-questioning approach compared to the free recall because reminiscence occurred. Reminiscence refers to the recall of information at succeeding recall tests which was not reported previously (Poole & White, 1991; see Payne, 1987 for a literature review). The phenomenon of reminiscence was found in several previous experimental studies using different participant populations (children with and without LD and adults), stimuli and delays (Cederborg, La Rooy & Lamb, 2008; La Rooy, Pipe & Murray, 2005; Poole & White, 1991; Turtle & Yuille, 1994). During the description task in Study 2, the cued recall always followed the free recall and

could therefore be regarded as a repeated recall/description attempt, which led to the retrieval of additional information. A possible explanation for the obtained reminiscence effect in the present study may be a failure of participants to spontaneously retrieve specific information about all the facial features at the first recall attempt. During the successive cued-questioning approach, specific questions were asked about all the individual facial features, thereby guiding participants' attention towards each feature, resulting in a more detailed description of the target faces.

The finding that people with mLD perform at a lower level in general than individuals without LD, but nevertheless profit from similar circumstances, provides evidence for the developmental approach to LD (Hodapp, Burack & Zigler, 1998). As mentioned briefly in Chapter 3, advocates of the developmental approach assume that people with LD proceed through the same developmental sequences as normally developed individuals, but at a slower rate. Moreover, the developmental approach states that children with LD should show similar performance to normally developed children matched on MA. Support for the developmental approach derives from studies by Henry and Gudjonsson (1999) and Michel et al. (2000), who found that children with LD do at least show recall performance for observed events which is similar to that of children matched on MA. In contrast to the developmental approach, the difference approach assumes that LD is the result of either deficits or differences in specific cognitive processes. Advocates of the difference approach argue that general developmental principles do not apply to individuals with LD (Bennet-Gates & Zigler, 1998). Moreover, they postulate that even when matched on MA, LD individuals will still display differences in performance due to intrinsic differences, which are independent of IQ. Although, it is not possible to provide explicit evidence

for either of these theoretical frameworks on the basis of the present research, since the experimental studies included only a CA matched control group, the current findings favour the developmental approach more.

The following practical implications derive from these findings: First, police officers or other professionals who engage in interviewing people with LD should start with a very general open-ended question format before proceeding to a more specific questioning style. However, additional specific questions might be unavoidable, because of the paucity of information, particularly facial information, provided by individuals with LD. This hierarchical questioning style is in agreement with the one recommended in the *Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses, and using special measures* document produced in 2007 (CJS, 2007) and the recently updated version of *Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children* (Home Office, 2002) (updated in 2008). Moreover, interviewers need to be aware that this questioning style might result in a quantity-quality trade off. Most importantly, the findings suggest that measures which are ideal for people without LD, may to a similar degree, also be beneficial for LD individuals.

11.3.2 The ability of people with mLD to construct facial composites

The inferior face recognition and description skills of people with mLD also manifest themselves during the construction of facial composites with contemporary composite systems. Study 3 and 5 described in this thesis demonstrated that people with mLD constructed significantly poorer composites than their non-LD counterparts. This finding is not particularly surprising, since the construction of facial composites requires the witness to engage in face recognition, recall and description (Pike et al.,

2000). Notwithstanding their poor composite construction abilities, some of their composites were good enough to act as reference for accurate target identification during subsequent matching tasks. Of particular interest is the reason why the composites by LD individuals were poorer than those created by control participants. The results obtained from Study 3 and 5 provide some potential explanations.

First, in line with the findings revealed in Study 2, individuals with mLD provide considerably less verbal information about the target faces during the CIs prior to the composite construction process with E-FIT. This is in accordance with previous research, investigating the effectiveness of the CI with LD witnesses. A recurring finding is that although the CI elicits more information from both individuals with and without LD about to-be-remembered events than a SI, the recall of people with LD is nevertheless poorer than the one provided by their non-LD peers (Brown & Geiselman, 1999; Milne & Bull, 1996; Milne et al., 1999; Robinson & McGuire, 2006). As reviewed in Chapter 4, a detailed verbal description of the perpetrator's face can be considered as a prerequisite for accurate composite construction, especially with the more featurally based composite systems, such as E-FIT. The verbal description provided by the witness determines the initial composite quality, which can subsequently be enhanced by making changes to the individual features. It appears logical that it is much harder to improve a poor initial starting point, than to enhance an already passable one. The correlational analysis conducted as part of Study 3 offers support for this argument and reveals a significant positive correlation between composites' likeness rating scores and the number of facial information provided during the CI. Further evidence, for the argument that the lack of verbal facial information provided during the CIs was at least partially responsible for the poor E-FIT composite likeness, derives from the finding that mLD participants

created significantly better composites with a system not requiring a verbal facial description. Study 5 demonstrated that participants with mLD generated considerably better composites with the purely holistic system EvoFIT. Thus, a composite system which relies less on language abilities of the witness appears to be more suited to the needs of witnesses with mLD.

In addition to limited verbal descriptions of a target face, people with mLD also provided sparse non-verbal information, such as pointing to the monitor or to their own face to facilitate operator-witness communication in Study 3. Indeed Brace et al. (2006) argue that the interaction between the witness and the operator may play a crucial role in the construction of an accurate composite image. They found that composites were rated as a significantly better likenesses when they were created by the operator alone than together with a witness. It could be argued that providing the operator with additional gestures (e.g. indicating how long the nose of the target face was by pointing to ones own nose, or pointing to the monitor to show how short the hair needs to be cut) facilitated the understanding of the operator and reduced any ambiguities. To reiterate, describing a face can be regarded as a very demanding and subjective task; what may be described as a big nose by one person may be viewed as small to another. Using additional gestures may have complemented the verbal facial descriptions and therefore considerably influenced resultant composite quality. Evidence for this assumption derives again from correlational analyses reported in Study 3, which revealed that composite likeness was significantly related to the amount of non-verbal behaviours displayed by participants. Thus, mLD participants provided the operator with very sparse verbal as well as non-verbal information about the target face. This lack in communication might have, to some extent, detrimentally influenced the quality of the resulting composites.

Furthermore, individuals with mLD spend significantly less time with the construction of the composite images than members of the control group. This was true for composite constructions with both E-FIT and EvoFIT. Potential reasons for this finding may be that mLD individuals wanted to change fewer facial features than control participants. They requested to see fewer alternative feature exemplars during the composite construction with E-FIT. Moreover, they were more easily satisfied with the applied changes and the resultant composite image in general, which led to the potentially premature termination of the composite construction. These findings fit with the expressed opinions of one of the E-FIT operators surveyed within our initial survey study. This operator indicated that the E-FIT system might be particularly unsuitable for witnesses with LD because of their reputation to be ‘people pleasers’. Consequently they become more reluctant to make changes to the composite image created. In addition, they are in line with previous research findings regarding mLD individuals’ proneness to acquiesce, the heightened tendency to answer specific questions in an affirmative way (Clare & Gudjonsson, 1993; Heal & Sigelman, 1995; Rapley & Antaki, 1996; Sigelman, Budd, Spanhel & Schoenrock, 1981). Thus, the reluctance of mLD individuals to make changes to the composite image and their easy to satisfy nature may have contributed to the circumstance that they spent less time with the composite construction than control individuals. This may have consequently resulted in poorer composite quality. Further supportive evidence for this claim derives again from correlational analyses in Study 3 and 5, which revealed that engaging longer with the composite construction led to superior likenesses.

In contrast to the composite construction with E-FIT, during the EvoFIT construction it is not entirely up to the witness to determine when a reasonable likeness emerged and hence the composite construction is completed. Even though the

final decision to bring the composite construction to an end lies with the witness, prior to this decision every witness is required to make use of a predetermined series of construction and enhancement techniques, i.e. the selection and breeding phase (which is run through twice), the holistic tool phase and the shape tool phase. The completion of each of these three phases is essential before the composite construction process with EvoFIT can be finished. The finding that individuals with mLD generated better likenesses with EvoFIT than with E-FIT suggests that a system which provides mLD witnesses with less choices but with more time, and multiple ways to work on the composite image may result eventually in superior likenesses.

The above cited findings suggest that the ability to create an accurate composite image depends not only on the capacities of the witness but also on the utilised system and whether this is concordant with the competencies and needs of the particular individual. The findings comprised in this thesis imply that EvoFIT should be regarded as the most suitable facial composite system for individuals with mLD. The system is less dependent on language and provides witnesses with a more structured composite construction approach, helping to support the language deficiencies held by mLD individuals and to counteract their increased tendency to acquiesce.

In addition, the results of the control group in Study 5 suggest that the efficiency of the utilised composite system also depends on situational factors, such as the exposure duration of the perpetrator's face and the delay between the witnessed incident and the following composite construction. The finding that exposure duration and delay can impact considerably upon subsequent face recognition abilities is not new and has been repeatedly demonstrated in previous applied face recognition studies (Krouse, 1981; Laughery et al., 1971; Reynolds & Pedzdek, 1992; Walker-

Smith, 1978; see Chapter 4 section 4.2.2 for a brief review). However, what is novel is the idea that these two situational factors can influence whether a more featural or holistic composite system produces better likenesses. As discussed in more detail in Chapter 10, Study 5 found that control participants do not benefit from the advantages of the EvoFIT system from which mLD individuals profited. Instead they generated superior composites with the more featural E-FIT system. At first glance, this finding seems to contradict recently conducted research, which revealed that composites created with EvoFIT resulted in better resemblances than those generated with more featural systems, such as E-FIT or ProFIT (Frowd et al., 2007; 2010). However, the present finding supports earlier studies (Frowd et al., 2005a; 2005b), which obtained no superior performance with EvoFIT. It could be argued that differences in methodology were responsible for the conflicting findings. Asking participants to view a target face for one minute and informing them that they will be later required to construct a facial composite of it, might have lead participants to engage in a more featural encoding approach, i.e. scanning the face from top to bottom and trying to memorise every single feature as accurately as possible. The idea that people can change their facial encoding strategy depending on the task demands has been demonstrated by Hole (1994) and Farah et al., (1995). As described in more detail during the discussion in Chapter 10, Hole (1994) revealed that decreasing the exposure time from two seconds to 80 milliseconds encouraged people to engage in more holistic face processing, as evidenced by the revealed face inversion effect. On the other hand, Farah et al. (1995) demonstrated that instructing participants to engage in a featural encoding approach can prevent them from displaying the face inversion effect. Thus, depending on the exposure duration of the target face and the task

instructions at hand people may engage in either featural or holistic facial encoding strategies.

Beside the applied target exposure duration, the delay between the presentation of the target and the subsequent composite construction may impact later face recall, specifically peoples' ability to recall individual features of the target face. In line with the fuzzy-trace framework (Brainerd & Gordon, 1994; Brainerd & Reyna, 1998; 2002; 2004; Reyna and Farrell, 1997), it could be argued that an increase in delay is attended by an inaccessibility of a detailed memory about the individual facial features. To recap, FTT makes the following assumptions: First, memory is not unitary, but can be divided into gist and verbatim representation. Although these two types of information are encoded in parallel, gist and verbatim representations are stored separately. Consequently, gist and verbatim representations can be elicited independently depending on the specific cues utilised (Reyna & Farrell, 1997). Second, compared to gist representations, verbatim representations are more susceptible to interference and they become more rapidly inaccessible when time passes (Reyna & Farrell, 1997). This has been demonstrated by numerous studies using different types of stimuli, such as word lists (Roediger & McDermott, 1995; Toggia, Neuschatz & Goodwin, 1999) numerical (Brainerd & Gordon, 1994) and pictorial information (Gernsbacher, 1985). FTT provides potential explanations for a variety of memory errors and has gained more and more popularity in the eyewitness research domain (Koriat, Goldsmith & Pansky, 2000; Wright & Loftus, 1998).

Relating the above research to memory for faces, which can be regarded as pictorial information, it could be reasoned that with increased delays between presentation and test our memory for individual facial features declines, or perhaps becomes less accessible. However, the holistic memory for the face and its feature

configurations, a more general impression, remains more or less unaffected. Depending on the type of facial representation (featural or holistic) in memory at the time of recall, people will produce better composites with systems congruent with their stored representation. It follows that after longer delays (e.g. four days) people generate superior composites with holistic systems while after shorter delays (two to three hours) better composites are generated with more featural systems. However, this remains an empirical question to be addressed in future research, and it is acknowledged that the studies comprised in this thesis only provide indirect evidence for this hypothesis. Further research would be required to manipulate the delay and the type of composite system utilised to verify this assumption. Despite EvoFIT being potentially incongruent with participants' memory representations in the present experimental studies, it nevertheless appeared to be more suitable for individuals with mLD than E-FIT. It could be reasoned that with respect to individuals with mLD, the non-verbal nature of the EvoFIT system has outweighed the facial representation incongruity.

11.3.3 Individual differences in face recognition and description

The findings of Study 2, 3, and 5 suggest that there may be differences in people's ability to accurately recognise and describe unfamiliar faces, and their ability to construct facial composites. Although face recognition is a well-studied area in psychology, there is only limited research available that investigates individual differences. This is surprising, since research regarding individual variation could shed further light on theoretical as well as more applied aspects of face processing. For example, by exploring associations between peoples' performance on tasks involving face recognition and those addressing face description, it could be established, whether these processes are related or independent of each other.

Furthermore, the study of individual differences in face recognition may also play an important role in more applied settings, such as during the recruitment of police and security personnel, who routinely need to match faces on the basis of CCTV footage or photographic IDs.

Ellis et al. (1975) were one of the first researchers who addressed this issue. They revealed that people show variability in their composite construction abilities with Photo-FIT. Those individuals who performed better during a composite reconstruction task, involving matching individual facial features, performed also superior during the composite construction from memory, a task which involves recognition memory. Thus, individuals' face matching abilities appear to be correlated with their face recognition skills. Furthermore, although not actually tested, Ellis et al. (1975) postulated that factors such as age, occupation and intellectual functioning may influence an individual's ability to create high quality composite images.

Evidence that intellectual functioning may indeed play an important role during tasks involving face recognition and description, derives from Study 2 in this thesis, which revealed that mLD individuals' WASI scores were significantly correlated with their performance on the main face recognition task and the amount and accuracy of the facial information provided during the face description tasks. Thus, mLD individuals with higher intellectual functioning perform better at tasks involving accurately recognising and describing unfamiliar faces. Moreover, Study 5 revealed significant correlations between mLD individuals' WASI scores and their performance on the morph task, which can be considered as a face matching task. In view of these findings, it can be concluded that individual differences in intellectual functioning can impact face perception and processing, encompassing matching, recognition, and memory for unfamiliar faces.

It should be noted that people with mLD in general displayed great individual differences in their performance on all of the face recognition and description tasks included in this thesis. Further, the variation in performance of the mLD participant group was considerably larger than in the control sample as evidenced by the significant Leven's tests on a substantial amount of tasks. Most importantly, it should be noted that some of the mLD individuals exhibited performance at the same level as members of the control group. This finding supports the proposition that witnesses with mLD can perform at a similar level as witnesses without LD.

Further support for individual variations in face processing comes from a recently conducted study by Burton, White and McNeill (2010). Substantial individual differences were revealed in performance on the Glasgow Face Matching Test (GFMT). The GFMT comprises 168 pairs of faces, of which half are same-face pairs and half different-face pairs. The test requires individuals to indicate whether the presented face pairs are identical or different. Burton et al. (2010) found that overall accuracy ranged from 62% to 100%, which can be considered as a rather wide range given the simplicity of the task. Additionally, Burton et al. (2010) found significant correlations between participants' GFMT scores and their performance on an old/new face recognition task. This is in line with the results obtained during the EvoFIT study (Study 5) in this thesis, which revealed significant correlations between participants' performance on the morph task (a face matching task) and their accurate construction of facial composites with EvoFIT, involving recognition memory. Thus, it seems as if the processes involved in matching, remembering, and recognising unfamiliar faces are to some extent related and dependent on each other. This suggests, that it may be possible to predict an individual's performance on one face processing task on the basis of his/her performance on another face processing task.

This inference bears important practical implications for legal settings, in particularly the construction of facial composites. The present findings suggest that easy and quick to deliver face matching or recognition tasks (such as the GFMT or the Cambridge Face Memory test (Duchaine & Nakayama, 2006)) could be used by the police to train witnesses to improve their performance during the construction of facial composites. Future research should address these possibilities, which may be particularly valuable in cases where the witness has a LD. Moreover, it should be investigated whether such tasks may be also suitable for screening appropriate witnesses.

11.4 Methodological limitations

It is acknowledged that the current studies are not free of limitations, and there are a few methodological shortcomings which should receive attention. These are considered in more detail below.

As with most research investigating the performance of eyewitnesses it could be argued that several aspects of the study lack ecological validity. First, the experimental environment was very artificial and conducive. Participants with mLD were tested at day care centres to which they attended on a daily basis. Thus, they were tested in a familiar environment in which they felt very comfortable and secure. In addition, all participants were willing and happy to take part and factors that may have led to feelings of distress, particularly in mLD participants, were avoided at all cost. Thus the levels of stress or anxiety within these experimental studies were probably not akin to those that may be present if these individuals were involved in real police interrogation scenarios. However, the experimental procedures utilised

were the only viable option for the present participant group due to their vulnerability and the involved ethical considerations.

A second deviation from real life is the use of static target photographs as stimuli material. This rather convenient procedure was applied during various previous composite construction studies (Brace et al., 2006; Cutler et al., 1988; Davies, 2000; Frowd et al., 2004; 2005a; 2005b; 2007; 2010; Kovera et al., 1997). It could be argued that their employment may have influenced the obtained results and makes them less able to be generalised. However, a meta-analysis conducted by Shapiro and Penrod (1986) evaluated 13 face perception studies and found only minor changes in participants' behaviour between studies involving live or video-taped stimuli and those presenting static photographs. This indicates that a static stimulus is unlikely to have affected the composite quality in any detrimental way. On the other hand, it may have actually aided composite construction. According to Van Koppen and Lochun (1997), in the real world witnesses often have only limited views on the perpetrator's face due to situational factors, such as distance, movement and/or disguise. Thus, inspecting a full-face static target photograph for one minute prior to the composite construction can be regarded as a very idealistic condition, which is rare maybe even nonexistent in a real-life eyewitness situation. Therefore it may be desirable for future research aiming to replicate the present findings to utilise more ecologically valid stimulus material such as real faces in a live encounter or in video clips.

With respect to the stimulus material, a further deviation from the real world is the use of unfamiliar faces during all experimental studies comprised in this thesis. It could be argued that using unfamiliar faces in composite construction research is once again not ecologically valid. In reality witnesses who are asked to create a composite

by the police are usually unfamiliar with the perpetrator. However, the people who subsequently identify the perpetrator on the basis of the composite image are usually familiar with him/her (friends, neighbours, relatives, colleagues). Consequently, the most ecologically valid procedure would be to choose famous faces as targets, which are unfamiliar to participants who construct the composite image but familiar to those who evaluate it subsequently. Unfortunately, this procedure is associated with a high drop-out rate because participants who recognise the famous targets prior to the composite construction need to be dismissed. Given that during the present experimental studies participants with mLD participated and the associated difficulty of getting access to this population contributed to the pragmatic decision to use unfamiliar faces as targets.

Instead of a naming task, a matching task and a likeness rating were employed as means of evaluation during Study 3 and 5. Frowd et al. (2007) argue that a matching task and a likeness rating may be not the most suitable evaluation instruments when testing a holistic composite system. Furthermore, the likeness rating can be considered as a rather subjective task. However, given the fact that on the basis of the chosen stimuli material it was not possible to apply a spontaneous naming task, it was considered best to use two evaluations tasks which measured the utility of the composite system in different ways. In Study 3 and 5 the results of the matching task and the likeness rating were in agreement with each other, providing evidence that the two tasks indeed tap into related processes. Furthermore, as mentioned by Paine (2004), a matching task does not give any information about relative differences, whereas the likeness rating does, thereby complementing the results of the matching task. Nevertheless, we agree that in an ideal situation a spontaneous naming task should have been used as a means of evaluation but due to pragmatic reasons outlined

above this was not possible. Overall, we agree with Paine's (2004) suggestion that multiple evaluation tasks should be preferably used during composite construction research.

A three hour delay between the presentation of the target face and the construction of the composite was applied during the present experimental studies to add ecological validity (similar delays were used in studies conducted by Frowd et al., 2005a; 2005b). However, such a delay may be regarded as rather short compared to those employed by other researchers. For example, Frowd et al., (2007; 2010) used a considerably longer delay of two days, although the ACPO Working group for facial Identification guidelines (2007) state that whenever possible, witnesses should be contacted to create a facial composite as soon as possible after the incident has happened. However, due to the nature of the investigation this is often not possible. Paine (2004) argues that sometimes delays can range from two days to six months. Due to pragmatic reasons, for example the busy daily activity schedules of mLD individuals and their heightened need to follow their usual routines at the day care centres, a two to three hours delay was the maximum we could employ.

Another limitation of all experiments included in this thesis is the use of students and members of staff as control participants. In general, students do not constitute an ecologically valid and representative sample population because of their homogeneous nature and their presumably above-average intellectual functioning. This sampling bias might have even larger effects when the experimental group of interest includes individuals with mLD. It could be argued that a sample consisting exclusively of university students and members of staff with higher than average intellectual functioning led to an exaggeration of the revealed group differences in performance on the experimental tasks. Therefore, future studies aiming to replicate

the findings obtained during this thesis should use a more ecologically valid control group, including participants from a wide variety of social backgrounds.

Overall, due to the special group of participants used in the current experimental studies and other factors inherent to laboratory based research itself, it was not at all times possible to opt for the most ecologically valid procedures. However, we have tried to create an experimental setting that produced findings which can be generalised to eyewitness situations that police officers would face in their dealings with mLD individuals. Future research is required to replicate the revealed findings under more ecologically valid conditions.

11.5 Concluding remarks

In conclusion, the obtained findings in this thesis have important practical implications, particularly for the police and other practitioners working in the legal field. Overall, witnesses with mLD should not be excluded from standard police procedures. This general conclusion is of utmost importance, particularly because in many cases individuals with LD are the only witnesses to a crime (Milne & Bull, 2001), including hate crimes and as well as victimisation against other people with LD. Despite the performance of these individuals being considerably poorer on a variety of face recognition and description tasks during the present research, their performance, importantly, was above the chance level, thus demonstrating capability. Furthermore, they showed variability in their performance dependent upon the measures introduced. Thus witnesses with LD, when questioned in an appropriate manner taking their disability into account, can provide accurate perpetrator descriptions, which may aid in the search and the subsequent successful apprehension of the offender/s.

Besides these rather general conclusions, the present thesis provides also more specific recommendations for composite operators on how to effectively generate facial composites with LD witnesses. Particularly, if there are two witnesses to a crime, one with mLD and one without LD, and the police wants to create a facial composite image of the perpetrators' face with them, it is probably more reasonable on the basis of the current findings, to rely on the composite created by the witness without LD. However, if there is only one witness, and this witness happens to have mLD, the composite should be created with EvoFIT rather than with E-FIT. This advice is invaluable, especially because specific recommendations for police operators were so far lacking in guideline documents, such as in the *Facial Identification Guidance 2009* (ACPO & NPIA, 2009).

It is desirable and essential to strengthen the present recommendations through future research replicating the findings comprised in this thesis. This will hopefully lead to alterations and adjustments of current guidelines and maybe even stimulate legislative changes that contribute to a more adequate and research-led treatment of individuals with LD by the criminal justice system. Furthermore, future research in the eyewitness domain, particular in the vulnerable witness area, should focus towards a positive psychology, i.e. investigating ways to improve performance and to solve problems rather than identifying limitations.

References

- ABM United Kingdom Ltd. (2008). *Sex attacker convicted using new facial composite technology, EvoFIT*. ABM United Kingdom Ltd.
- ACPO & Home Office. (2002). *Vulnerable Witnesses: A Police Service Guide*. London: ACPO & Home Office.
- ACPO (2003). *National working practice in facial imaging*: ACPO.
- ACPO (2007). *Working group for facial Identification guidelines*. ACPO
- ACPO, & NPIA (2009). *Facial Identification Guidance 2009*. ACPO.
- Agnew, S. E. & Powell, M. B. (2004). The effect of intellectual disability on children's recall of an event across different question types. *Law and Human Behavior, 28*, 273-294.
- American Psychiatric Association (1995). *Diagnostic and Statistical Manual of Mental Disorders (4th Ed)*. Washington D.C.: American Psychiatric Association.
- Aschermann, E., Dannenberg, U., & Schulz, A. (1998). Photographs as retrieval cues for children. *Applied Cognitive Psychology, 12*, 55-66.
- Ashton, G. R. & Ward, A. D. (1992). *Mental Handicap and the Law*. London: Sweet & Maxwell.
- Aspley Limited. (2004). *E-FIT. Net User Manual*. Hatfield: Apley Limited.
- BILD (2006). Factsheet: Learning disabilities. Retrieved from <http://www.bild.org.uk/05factsheets.htm>
- Bahrack, H. P., Bahrack, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: a cross-sectional approach. *Journal of Experimental Psychology: General, 104*(1), 54-75.
- Bailey, A., Barr, O., & Bunting, B. (2001). Police attitudes toward people with intellectual disabilities: an evaluation of awareness training. *Journal of Intellectual Disability Research, 45*(4), 344-350.
- Belmont, J. M., & Butterfield, E. C. (1971). Learning strategies as determinants of memory deficiencies. *Cognitive Psychology, 2*(4), 411-420.
- Bennet-Gates, D., & Zigler, E. (1998). Resolving the developmental-difference debate: An evaluation of the triarchic and system theory models. In J. A. Burack, R. M. Hodapp, & E. Zigler (Eds.), *Handbook of mental retardation and development* (pp. 3-19). Cambridge: Cambridge University Press.
- Benson, P. J., David, I., & Perrett, I. (1994). Visual processing of facial distinctiveness. *Perception, 23*(1), 75-93.
- Benton, A. L., & Van Allen, M. W. (1972). Prosopagnosia and facial discrimination. *Journal of Neurological Sciences, 15*(2), 167-172.
- Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect. Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science, 18*(8), 706-712.

- Brooks, B. M., Rose, F. D. Attree, & Elliot-Square, A. (2002). An evaluation of the efficacy of training people with learning disabilities in a virtual environment. *Disability and Rehabilitation*, 24(11-12), 622-626.
- Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-rational identification. *Personality and Social Psychology Bulletin*, 15(1), 19-25.
- Bower, G. H. (1967). A multicomponent- theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation Vol. 1* (pp. 229-325). New York & London: Academic Press.
- Brace, N. A., Pike, G. E., Allen, P., & Kemp, R. I. (2006). Identifying composites of famous faces: Investigating memory, language and system issues. *Psychology, Crime & Law*, 12(4), 351-366.
- Brace, N. A., Pike, G. E., & Turner, J. A. (2008). Holistic facial composite systems: are they compatible with witness recall? *Cognitive Technology Journal*, 13(2), 42-49.
- Bradshaw, J. (2001). Complexity of staff communication and reported level of understanding skills in adults with intellectual disability. *Journal of Intellectual Disability Research*, 45(3), 233-243.
- Brainerd, C. J., & Gordon, L. L. (1994). Development of verbatim and gist memory for numbers. *Developmental Psychology*, 30(2), 163-177.
- Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology*, 71, 81-129.
- Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, 2(5), 164-169.
- Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. *Developmental Review*, 24, 396-439.
- British Psychological Society (2000). *Learning disability: Definition and Contexts*. Professional Affairs Board of The British Psychology Society, Leicester.
- Brown, A. L., & Barclay, C. R. (1976). The effects of training specific mnemonics on the metamnemonic efficiency of retarded children. *Child Development*, 47, 71-80.
- Brown, C. L. & Geiselman, R. E. (1990). Eyewitness testimony of mentally retarded: Effect of the cognitive interview. *Journal of Police and Criminal Psychology*, 6, 14-22.
- Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, 73, 105-116.
- Bruce, V. (1988). *Recognising faces*. Hove, UK: Lawrence Erlbaum Associates, Publishers.
- Bruce, V., Burton, M., & Hancock, P. (2007). Remembering faces. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness Psychology. Volume II Memory for People* (pp. 87-100). Mahwah, New Jersey: Lawrence Erlbaum Association, Publishers.

- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5(4), 339-360.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305-327.
- Bruce, V., & Young, A. (1998). *In the eye of the beholder. The science of face perception*. Oxford: Oxford University Press.
- Burack, J., Hodapp, R., & Zigler, E. (1998). *Handbook of Mental Retardation and Development*. New York & Melbourne: Cambridge University Press.
- Burton, M., Evans, R., & Sanders, A. (2006). *Are special measures for vulnerable and intimidated witnesses working? Evidence from the criminal justice agencies*. London: Home Office.
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, 42(1), 286-291.
- Butterfield, E. C., Wambold, C., & Belmont, J. M. (1973). On the theory and practice of improving short-term memory. *American Journal of Mental Deficiency*, 77(5), 654-659.
- Cederborg, A., & Lamb, M. (2008). Interviewing alleged victims with intellectual disabilities. *Journal of Intellectual Disability Research*, 52(1), 49-58.
- Cederborg, A., La Rooy, D., & Lamb, M. (2008). Repeated interviews with children who have intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, 21, 103-113.
- Chance, J. E., & Goldstein, A. G. (1996). The other-race effect and eyewitness identification. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 87-115). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Christie, D. F. M., & Ellis, H. D. (1981). Photofit constructions versus verbal descriptions of faces. *Journal of Applied Psychology*, 66(3), 358-363.
- CJS (2007). *Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses, and using special measures*. London: Criminal Justice System.
- Clare, I. C., & Gudjonsson, G. H. (1993). Interrogative suggestibility, confabulation, and acquiescence in people with mild learning disabilities (mental handicap): implications for reliability during police interrogations. *British Journal of Clinical Psychology*, 32(3), 295-301.
- Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(4), 739-755.
- Clifford, B., & George, R. (1996). A field evaluation of training in three methods of witness/victim investigative interviewing. *Psychology, Crime & Law*, 2(3), 231-248.
- Cooke, P., & Davies, G. (2001). Achieving best evidence from witnesses with learning disabilities: new guidance. *British Journal of Learning Disabilities*, 29(3), 84-87.

- Cutler, B. L., Stocklein, C. J., & Penrod, S. D. (1988). Empirical examination of a computerized facial composite production system. *Forensic Reports, 1*(3), 207-218.
- Dando, C. J., & Milne, R. (2009). The cognitive interview. In R. N. Kocsis (ED.), *Applied criminal psychology, a guide to forensic behavioural sciences*. Springfield, IL: Charles.C. Thomas.
- Davies, L. A. (1995). *People with mental retardation in the criminal justice system*. Retrieved November the 16th, 2007, from <http://www.thearc.org>.
- Davies, L. A. (2005). People with intellectual disabilities in the criminal justice system: Victims & suspects. *The Arc*.
- Davies, G., & Christie, D. (1982). Face recall: An examination of some factors limiting composite production accuracy. *Journal of Applied Psychology, 67*(1), 103-109.
- Davies, G., Ellis, H., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology, 63*(2), 180-187.
- Davies, G., & Griffiths, L. (2008). Eyewitness identification and the English courts: A century of trial and error. *Psychiatry, Psychology and Law, 15*(3), 435-449.
- Davies, G., & Milne, A. (1982). Recognizing faces in and out of context. *Current Psychological Research, 2*(4), 235-246.
- Davies, G. M., & Valentine, T. (2006). Facial composites: Forensic utility and psychological research. In R. C. C. Lindsay, D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology* (pp. 54-96). Mahawa, NJ: Erlbaum.
- Davies, G., Van der Willik, P., & Morrison, L. J. (2000). Facial composite production: A comparison of mechanical and computer-driven systems. *Journal of Applied Psychology, 85*(1), 119-124.
- Dent, H. (1986). An experimental study of the effectiveness of different techniques of questioning mentally handicapped child witnesses. *British Journal of clinical Psychology, 25*, 13-17.
- Department of Health (2001). *Valuing People: A new strategy for learning disability for the 21st century*. London: Stationery Office.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology, 115*(2), 107-117.
- Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effects: Why descriptions impair face recognition. *Memory & Cognition, 25*(2), 129-139.
- Dobson, E., & Rust, J. O. (1994). Memory for objects and faces by the mentally retarded and non retarded. *Journal of Psychology, 128*(3), 315-322.
- Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic subjects. *Neuropsychologia, 44*(4), 576-585.

- Edwards, L. (2005). Switching off the surveillance society? Legal regulation of CCTV in the United Kingdom. In S. Nouwt, B. R. de Vries, & C. Prins (Eds), *Reasonable expectations of privacy? Eleven country reports on camera surveillance and workplace privacy* (pp.91-114). The Hague, NL: T.M.C Asser Press.
- Ellis, H. D. (1975). Recognizing faces. *British Journal of Psychology*, 66(4), 409-426.
- Ellis, H. D. (1986). Face recall: A psychological perspective. *Applied Cognitive Psychology*, 5(4), 189-196.
- Ellis, H. D., Davies, G. M., & Shepherd, J. W. (1978). A critical examination of the Photo-Fit system for recalling faces. *Ergonomics*, 21(4), 297-307.
- Ellis, H. D., & Shepherd, J. W. (1992). Face memory-Theory and practice. In M. Gruneberg, & P. Morris (Eds.), *Aspects of memory* (pp. 51-85). London, Routledge.
- Ellis, H. D., Sheperd, J., & Davies, G. (1975). An investigation of the use of the Photo-Fit technique for recalling faces. *British Journal of Psychology*, 66(1), 29-37.
- Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, 8(4), 431-439.
- Ellis, H., Shepherd, J., & Davies, G. (1980). The deterioration of verbal descriptions over different delay intervals. *Journal of Police Science and Administration*, 8, 101-106.
- Ellison, L. (2001). *The adversarial process and the vulnerable witness*. Oxford & New York: Oxford University Press.
- Emerson, E. (2001). *Challenging behaviour: analysis and intervention in people with severe intellectual disabilities*. Cambridge: Cambridge University Press.
- Emerson, E., & Hatton, C. (1996) Deinstitutionalization in the UK and Ireland: Outcomes for service users. *Journal of Intellectual & Developmental Disability*, 21, 17-37.
- Emerson E., & Hatton C. (2008). *People with learning disabilities in England*. Lancaster: Centre for Disability Research (CeDR): Lancaster University.
- Emerson, E., Hatton, C., Felce, D. & Murphy, G. (2001). *Learning disabilities: The fundamental facts*. The Mental Health Foundation, London.
- eSAY (2008). *Statistics release. Adults with learning disabilities implementation of 'The same as you?' Scotland 2008*. Retrieved August, 27th, 2009 from <http://www.sclld.org.uk>
- Ewbank, M. P., & Andrews, T. J. (2008). Differential sensitivity for viewpoint between familiar and unfamiliar faces in human visual cortex. *NeuroImage*, 40, 1857-1870.
- Farah, J. W., Drain, & Tanaka (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), 628-634.

- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1988). What is 'special' about face perception? *Psychological Review*, *105*(3), 482-498.
- Fisher, R. P., Geiselman, R., & Raymond, D. (1987). Critical analysis of police interview techniques. *Journal of Police Science and Administration*, *15*(3), 117-185.
- Frowd, C. D., Bruce, V., & Hancock, P. J. B. (2008). Helping the police to construct the face of a criminal. *The Psychologist*, *21*(8), 670-682.
- Frowd, C. D., Bruce, V., McIntyre, A., & Hancock, P. J. B. (2007). The relative importance of external and internal features of facial composites. *British Journal of Psychology*, *98*, 61-77.
- Frowd, C. D., Bruce, V., McIntyre, A. H., Ross, D., Fields, S., Plenderleith, Y., & Hancock, P. J. B. (2006). Implementing holistic dimensions for a facial composite system. *Journal of Multimedia*, *1*, 42-51.
- Frowd, C. D., Bruce, V., Ness, H., Bowie, L., Paterson, J., Thomson-Bogner, C., McIntyre, A. H., & Hancock, P. J. B. (2007). Parallel approaches to composite production: interfaces that behave contrary to expectation. *Ergonomics*, *50*, 562-585.
- Frowd, C. D., Bruce, V., Smith, A. J., & Hancock, P. J. B. (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied*, *14*(3), 276-287.
- Frowd, C. D., Carson, D., Ness, H., McQuinston-Surrett, D., Richardson, J., Baldwin, H., & Hancock, P. J. B. (2005b). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal & Criminological Psychology*, *101*(1), 63-81.
- Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S., & Hancock, P. J. B. (2005a). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law*, *11*(1), 33-52.
- Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perception*, *1*(1), 19-39.
- Frowd, C. D., Pitchford, M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., McIntyre, A. H., & Hancock, P. J. B. (2010). The psychology of face construction: Giving evolution a helping hand. *Applied Cognitive Psychology*, published online.
- Geiselman, E., Fisher, R. P., Firstenberg, I., Hutton, L. A., Sullivan, S. J., Avetissan, I. V., & Prosk, A. L. (1984). Enhancement of eyewitness memory - An empirical evaluation of the cognitive interview. *Journal of Police Science and Administration*, *12*(1), 74-80.
- Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, *17*(3), 324-363.
- Gibling, F., & Bennett, P. (1994). Artistic enhancement in the production of PhotoFIT likenesses: An examination of its effectiveness in leading to suspect identification. *Psychology, Crime & Law*, *1*(1), 93-100.

- Gibson, S., Pallares-Bejarano, A., & Solomon, C. (2003). Synthesis of photographic quality facial composite using evolutionary algorithms. In R. Harvey & J. A. Bangham (Eds.), *Proceedings of the British Machine Vision Conference* (pp. 221-230). London, UK: British Machine Vision Association.
- Gravetter, F. J., & Wallnau, L. B. (2007). *Statistics for the behavioral sciences (7th ed.)*. London: Thomson Wadsworth.
- Green, D. L., & Geiselman, R. E. (1989). Building composite facial images: Effects of feature saliency and delay of construction. *Journal of Applied Psychology, 74*(5), 714-721.
- Griffiths, A., & Milne, R. (2010). Application of the cognitive interview techniques as part of investigations. In C. A. Ireland & M. Fisher (Eds.), *Consulting and advising in forensic practice: Empirical and practical guidelines*. Chichester: Wiley.
- Gudjonsson, G. H. (1990). The relationship of intellectual skills to suggestibility, compliance and acquiescence. *Personality and Individual Differences, 11*(3), 227-231.
- Gudjonsson, G. H., & Henry, L. (2003). Child and adult witnesses with intellectual disabilities: The importance of suggestibility. *Legal and Criminological Psychology, 8*(2), 241-252.
- Hancock, P. J. B. (2000). Evolving faces from principal components. *Behavior Research Methods, Instruments & Computers, 32*(2), 327-333.
- Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar face. *Trends in Cognitive Sciences, 4*(9), 330-337.
- Hatton, C. & Emerson, E. (1996). Deinstitutionalization in the UK and Ireland: Outcomes for service users. *Journal of Intellectual & Developmental Disability, 21*, 17-37.
- Heal, L. W., & Sigelman, C. K. (1995). Response biases in interviews of individuals with limited mental abilities. *Journal of Intellectual Disability Research, 39*(4), 331-340.
- Henry, L. A. & Gudjonsson, G. H. (1999). Eyewitness memory and suggestibility in children with mental retardation. *American Journal on Mental Retardation, 104*(6), 491-508.
- Henry, L. A., & Gudjonsson, G. H. (2003). Eyewitness memory, suggestibility, and repeated recall sessions in children with mild and moderate intellectual disabilities. *Law and Human Behavior, 27*, 481-505.
- Henry, L. A., & Gudjonson, G. H. (2004). The effects of memory trace strength on eyewitness recall in children with and without intellectual disabilities. *Journal of Experimental Child Psychology, 89*, 53-71.
- Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 986-1004.
- Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*, 201-222.

- Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. (2004). *SPSS Explained*. London & New York: Routledge.
- Hodapp, R. M., Burack, J. A., & Zigler, E. (1998). Developmental approaches to mental retardation: A short introduction. In J. A. Burack, R. M. Hodapp, & E. Zigler (Eds.), *Handbook of mental retardation and development* (pp. 3-19). Cambridge: Cambridge University Press.
- Hogg, J. (2001). *Learning Disability in Tayside: The prevalence of people with learning disabilities and their associated needs in Angus, Dundee City & Perth & Kinross*. Dundee: White Top Research Unit, University of Dundee.
- Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, 23(1), 65-74.
- Home Office. (1998). *Speaking up for Justice*. London: Home Office.
- Home Office (2002). *Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children*. London: Her Majesty's Stationary Office.
- Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, 17(5), 577-596.
- Johnson, M. H., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition*. Oxford: Blackwell.
- Kebbel, M. R., & Hatton, C. (1999) People with mental retardation as witnesses in court: A review. *Mental Retardation*, 37(3), 179-187.
- Kebbel, M., Hatton, C., & Johnson, S. (2004). Witnesses with intellectual disabilities in court: What questions are asked and what influence do they have? *Legal and Criminological Psychology*, 9(1), 23-35.
- Koehn, C. E., & Fisher, R. P. (1997). Constructing facial composites with the Mac-a-Mug Pro system. *Psychology, Crime & Law*, 3(3), 209-218.
- Koehnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: a meta-analysis. *Psychology, Crime & Law*, 5(1,2), 3-27.
- Koriat, A., Goldsmith, M., & Pansky, A. (2000). Psychology of memory accuracy. *Annual Review of Psychology*, 51, 481-537.
- Kovera, M. B., Penrod, S. D., Pappas, C., & Thiel, D. L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology*, 82(2), 235-246.
- Krouse, F. L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology*, 66(5), 651-654.
- La Rooy, D., Pipe, M. E., & Murray, J. E. (2005). Reminiscence and hypermnesia in children's eyewitness memory. *Journal of Experimental Child Psychology*, 90(3), 235-254.
- La Rooy, D., Pipe, M., & Murray, J. (2007). Enhancing children's event recall after long delays. *Applied Cognitive Psychology*, 21, 1-17.
- Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixilation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology*, 15, 101-116.

- Laughery, K. R., Alexander, J. F., & Lane, A. B. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology, 55*(5), 477-483.
- Laughery, K., Duval, C., & Wogalter, M. (1986). *An analysis of procedures for generating facial images* (Mug File Project [Report No. UHMUG-2]). Houston, TX: University of Houston.
- Laughery, K. R., & Fowler, R. H. (1980). Sketch artist and Identi-kit procedures for recalling faces. *Journal of Applied Psychology, 65*(3), 307-316.
- Leichtman, M., & Ceci, S. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology, 31*(4), 568-578.
- Lindsay, D. S., Jack, P. C., & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology, 76*(4), 587-589.
- Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge: Harvard Press.
- MacLin, O. H., MacLin, M. K., & Malpass, R. S. (2001). Race, arousal, attention, exposure, and delay. An examination of factors moderating face recognition. *Psychology, Public Policy, and Law, 7*(1), 134-152.
- Malone, D. R., Morris, H. H., Kay, M. C., & Levin, H. S. (1982). Prosopagnosia: a double dissociation between the recognition of familiar and unfamiliar faces. *Journal of Neurosurgery, and Psychiatry, 45*, 820-822.
- Malpass, R. S., & Kravitz, J. (1969). Recognition of faces of own and other race. *Journal of Personality and Social Psychology, 13*(4), 330-334.
- Maurer, D., Le Grand, R., & Mondloch, C. (2002). The many faces of configural processing. *TRENDS in Cognitive Sciences, 8*(6), 255-260.
- McCartney, J. (1987). Mentally retarded and nonretarded subjects' long-term recognition memory. *American Journal of Mental Retardation, 92*(3), 312-317.
- McIntosh, R., Vaughn, S., & Zaragoza, N. (1991). A review of social interventions for students with learning disabilities. *Reports, 24*(8), 451-458.
- McQuiston-Surrett, D., Topp, L. D., & Malpass, R. S. (2006). Use of facial composite systems in US law enforcement agencies. *Psychology, Crime & Law, 12*(5), 505-517.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces. A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3-35.
- Meissner, C. A., Sporer, S L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness Psychology. Volume II Memory for People* (pp. 87-100). Mahwah, New Jersey: Lawrence Erlbaum Association, Publishers.
- Memon, A. (1997). The cognitive interview. In O. Hargie, A., *Handbook of Communication Skills* (pp. 451-470). London & New York: Routledge.
- Memon, A., & Bull, R. (1999). *Handbook of the psychology of interviewing*. New York & Chichester: John Wiley & Sons.

- Memon, A., & Koehnken, G. (1992). Helping interviewees to remember more: the cognitive interview. *Expert Evidence*, 1(2), 39-48.
- Memon, A., & Vartoukian, R. (1996). The effects of repeated questioning on young children's eyewitness testimony. *British Journal of Psychology*, 87, 403-415.
- Michel, M., Gordon, B., Ornstein, P., & Simpson, M. (2000). The ability of children with mental retardation to remember personal experiences: implications for testimony. *Journal of Clinical Child & Adolescent Psychology*, 29(3), 453-463.
- Milne, R. (1999). Interviewing children with learning disabilities. In Memon, A. & Bull, R., *Handbook of the Psychology of Interviewing* (pp. 165-180). West Sussex: John Wiley & Sons Ltd.
- Milne, R., & Bull, R. (1996). Interviewing children with mild learning disabilities with the cognitive interview. *Issues in Criminological and Legal Psychology*, 26, 44-51.
- Milne, R., & Bull, R. (2001). Interviewing witnesses with learning disabilities for legal purposes. *British Journal of Learning Disabilities*, 29, 93-97.
- Milne, R., & Bull, R. (2003). Does the cognitive interview help children to resist the effects of suggestive questioning? *Legal and Criminological Psychology*, 8(1), 21-38.
- Milne, B., & Bull, R. (2006) Interviewing Victims of Crime, Including Children and People with Intellectual Disabilities. In *Practical Psychology for forensic Investigations and Persecutions* ed. Kebbell, M. R. & Davies, G. M. pp.7-23. Chichester: John Wiley & Sons, Ltd.
- Milne, R., Clare, I. C. H., & Bull, R. (1999). Using the cognitive interview with adults with mild learning disabilities. *Psychology, Crime & Law*, 5(1-2), 81-99.
- Milne, R., Clare, I. C. H. & Bull, R. (2002). Interrogative suggestibility among witnesses with mild intellectual disabilities: The use of an adaptation of the GSS. *Journal of Applied Research in Intellectual Disabilities*, 15, 8-17.
- Moston, S. (1990). How children interpret and respond to questions: situational sources of suggestibility in eyewitness interviews. *Social Behaviour*, 5(3), 155-167.
- Narby, D. J., Cutler, B. L., & Penrod, S. D. (1996). The effects of witness, target, and situational factors on eyewitness identifications. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 87-115). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Nettelbeck, T. & Wilson, C. (2002). Personal vulnerability to victimisation of people with mental retardation. *Trauma, Violence, & Abuse*, 3, 289-306.
- O'Hara, J. (2001). A learning-disabled woman who had been raped: A multi-agency approach. *Journal Of The Royal Society Of Medicine*, 94, 245-246.
- O'Kelly, C., Kebbell, M., Hatton, C., & Johnson, S. (2003). Judicial intervention in court cases involving witnesses with and without learning disabilities. *Legal and Criminological Psychology*, 8(2), 229-240.
- Paine, C. B. (2004). *Children making faces: Enhancing children's facial recall and composite construction (Unpublished doctoral dissertation)*. The Open University, UK.

- Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory*, 3(4), 406-417.
- Payne, D. G. (1987). Hypermnnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, 101(1), 5-27.
- Peled, M., Iarocci, G., & Connolly, D. (2004) Eyewitness testimony and perceived credibility of youth with mild intellectual disabilities. *Journal of intellectual disability research* 48(7), 699-703.
- Penry, J. (1974). Photo-FIT. *Criminologist*, 9(31), 19-28.
- Penry, J., & Isobel, R. (1971). Looking at faces and remembering them: A guide to facial identification. London: Elek.
- Perlman, N. B., Ericson, K. I., Esses, V. M. & Isaacs, B. J. (1994). The developmentally handicapped witness. *Law and Human Behavior*, 18, 171-187.
- Petersilia, J. R. (2001). Crime victims with developmental disabilities A review essay. *Criminal Justice and Behavior*, 28, 655-694.
- Pike, G., Kemp, R. & Brace, N. (2000). The psychology of human face recognition. *IEE Electronics and Communication: Visual Biometrics*, 00/018, pp12/1-12/6.
- Pike, G. E, Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition*, 4(4), 409-437.
- Pipe, M., Lamb, M. E., Orbach, Y., & Esplin, P. W. (2004). Recent research on children's testimony about experienced and witnessed events. *Developmental review*, 24, 440-468.
- Plotnikoff J., & Woolfson, R. (2009). *Measuring Up? Evaluating implementation of government commitments to young witnesses in criminal proceedings*. NSPCC & Nuffield Foundation.
- Poole, D. A., & White, L. T. (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology*, 27, 975-986.
- Poole, D. A., & White, L. T. (1993). Two years later: Effects of question repetition and retention interval on the eyewitness testimony of children and adults. *Developmental Psychology*, 29(5), 844-853.
- Rapley, M., & Antaki, C. (1996). A conversation analysis of the 'acquiescenc' of people with learning disabilities. *Journal of Community & Applied Social Psychology*, 6(3), 207-227.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, 7(1), 1-75.
- Reyna, V. F., & Farrell, L. (1997). Theories of false memory in children and adults. *Learning and Individual Differences*, 9(2), 95-129.
- Reyna, V. F., & Kirnan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, 30(2), 178-191.
- Reyna, V. F., & Titcomb, A. C. (1996). Constraints on the suggestibility of eyewitness testimony: A fuzzy-trace theory analysis. In D. G. Payne & F. G.

- Conrad (Eds.), *Intersection in basic and applied memory research* (pp. 157-174). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Reynolds, J. K., & Pezdek, K. (1992). Face recognition memory: The effects of exposure duration and encoding instruction. *Applied Cognitive Psychology, 6*, 279-292.
- Robinson, J. & McGuire, J. (2006). Suggestibility and children with mild learning disabilities: The use of the cognitive interview. *Psychology, Crime & Law, 12*, 537-556.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology, 21*(4), 803-814.
- Scapinello, K. F., & Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition and pictorial stimuli. *Psychonomic Sciences, 21*, 329-330.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36-71.
- Scottish Government (2008). The population with learning disabilities in Scotland: A review of existing data. Retrieved August the 27th, 2009, from <http://cci.scot.nhs.uk>
- Severin, C. (2008). *No action over 'rape' of woman with learning difficulties*. Retrieved August, 14th, 2008 from <http://www.guradian.co.uk>.
- Shapiro, P., N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*(2), 139-156.
- Sharp, L. & Ross, M. (2008). *The Vulnerable Witnesses (Scotland) Act*. Dundee: Dundee University Press.
- Shepherd, J. W., & Ellis, H. D. (1996). Face recall-Methods and problems. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 87-115). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Shepherd, J. W., Ellis, H., & Davies, G. (1977). *Perceiving and remembering faces*. Report to the Home Office POL/73/1675/2411.
- Shriver, E. R., Young, S. G., Hugenberg, K., Bernstein, M. J., & Lanter, J. R. (2008). *Personality and Social Psychology Bulletin, 34*(2), 260-274.
- Sigelman, C. K., Budd, E. C., Spanhel, C. L., & Schoenrock, C. J. (1981). When in doubt, say yes: Acquiescence in interviews with mentally retarded persons. *Mental Retardation, 19*(2), 53-58.
- Sobsey, D. & Varnhagen, C. (1988). Sexual abuse and exploitation of people with disabilities. Final report, Alberta University Developmental Disabilities Centre, Edmonton, Canada, (Eric Document Reproduction Service No. 346620).
- Sporer, S. L. (1992). *I saw the man-but what did he look like? Person descriptions in an archival analysis of criminal cases*. Paper presented at the Biennial Meeting of the American Psychology-Law Society, San Diego, CA.

- Sporer, S. L. (2001). The cross-race effect. Beyond recognition of faces in the laboratory. *Psychology, Public Policy, and Law*, 7(1), 170-200.
- Sporer, S. L., Malpass, R. S., & Koehnken, G. (1996). *Psychological issues in eyewitness identification*. Mahwah, New Jersey: Lawrence Erlbaum Association, Publishers.
- Stobbs, G., & Kebbell, M. R. (2003) Jurors' perception of witnesses with intellectual disabilities and the influence of expert evidence. *Journal of Applied Research in Intellectual Disabilities* 16(2), 107-114.
- Swanson, H. L., Cooney, J. B. & O'Shaughnessy, T. E. (2004). Learning disabilities and memory. In B. Wong (3rd Eds.), *Learning about learning disabilities* (pp. 107-162). San Diego: Academic Press.
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 46A(2), 225-245.
- Tanaka, J. W., Farah, M. J., Peterson, M. A., & Rhodes, G. (2003). The holistic representation of faces. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 53-71). New York, NY US: Oxford University Press.
- Taylor, K. T. (2001). *Forensic art and illustration*. New York: CRC Press.
- The Association of Chief Police Officers (ACPO) National Working Practice in Facial Imaging document.
- Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception*, 9, 483-484.
- Tiddeman, B. P., Burt, M., & Perrett, D. I. (2001). Prototyping and Transforming Facial Textures for Perception Research. *IEEE Computer Graphics and Applications*, 21(5), 42-50.
- Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusionary memories: when more is less. *Memory*, 7(2), 233-256.
- Tredoux, C., Nunez, D., Oxtoby, O., & Prag, B. (2006). An evaluation of ID: an eigenface based construction system. *South African Computer Journal*, 37, 1-19.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352-373.
- Turtle, J. W., & Yuille, J. C. (1994). Lost but not forgotten details: Repeated Eyewitness recall leads to reminiscence but not hypermnesia. *Journal of Applied Psychology*, 79(2), 260-271.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43A(2), 161-204.
- Valentine, T. (1999). Face space models of face recognition. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives*, (pp. 83-115). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.
- Valentine, T., & Bruce, V. (1986a). Recognizing Familiar faces: The role of Distinctiveness and familiarity. *Canadian Journal of Psychology*, 40(3), 300-305.

- Valentine, T., & Bruce, V. (1986b). The effects of distinctiveness and classifying faces. *Perception, 15*, 525-535.
- Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effect of race and distinctiveness. *The Quarterly Journal of Experimental Psychology, 44A*(4), 674-703.
- Valentin-Hein, D. C., & Schwartz, L. D. (1993) Witness competency in people with mental retardation: Implication for prosecution of sexual abuse. *Sexuality and Disability, 11*(4), 287-294.
- Van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior, 21*(6), 661-685.
- VOICE, Respond, Mencap (2001). *Behind closed doors. Preventing sexual abuse against adults with a learning disability*. London: Mencap.
- Walker-Smith, G. J. (1978). The effects of delay and exposure duration in a face recognition task. *Perception & Psychophysics, 24*(1), 63-70.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York: The Psychological: Corporation.
- Welsh Assembly Government (2007). *Statement on policy and practice for adults with a learning disability*. Cardiff: Welsh Assembly Government.
- Wilson, C. & Brewer, N. (1992). The incidence of criminal victimisation on individuals with an intellectual disability. *Australian Psychologist, 27*, 114-117.
- World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.
- World Health Organization (2007) *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. Geneva: World Health Organisation.
- Wright, A. M., & Holliday, R. E. (2007). Enhancing the recall of young, young-old, and old-old adults with cognitive interviews. *Applied Cognitive Psychology, 21*(1), 19-43.
- Wright, D. B., & Loftus, E. F. (1998). How misinformation alters memories. *Journal of Experimental Child Psychology, 71*, 155-164.
- Wright, D. B., Memon, A., Skagerberg, E. M., & Gabbert, F. (2009). When eyewitnesses talk. *Current Directions in Psychological Science, 18*(3), 174-178.
- Yin, R. K. (1969). Looking at upside down faces. *Journal of Experimental Psychology, 81*, 141-145.
- Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception, 14*, 737-746.
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747-759.
- Young, A. W., Newcombe, F., de Haan, E. H. F., Small, M., & Hay, D. C. (1993). Face perception after brain injury. Selective impairments affecting identity and expression. *Brain, 116*, 941-959.

- Youth Justice and Criminal Evidence Act (1999). Retrieved April, 30th, 2010 from www.legislation.gov.uk
- Zarkadi, T., Wade, K. A., & Steward, N. (2009). Creating fair line-ups for suspects with distinctive features. *Psychological Science*, 20(12), 1448-1453.
- Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, 35(4)

Appendices

Appendix 1

A copy of the survey of police operators used in Study 1



UNIVERSITY
of
ABERTAY DUNDEE

Survey of Facial Composite Operators:

*Facial Composites and
Learning Disabilities*





POLICE SURVEY: FACIAL COMPOSITES AND LEARNING DISABILITIES

Dear sir or madam,

The purpose of this survey is to collect useful information about operators' experiences with witnesses with and without learning disabilities. To enhance the accuracy of composite systems and the performance of witnesses during the composite construction process, it is critical to identify the composite systems currently utilized by police stations, as well as methods of training and operators attitudes and treatment towards witnesses in general and witnesses with learning disabilities. Please take a few minutes to complete the questions below. Your answers will help to enhance police procedures for creating facial composites and to meet special requirements of witnesses with learning disabilities. Please feel free to give your own comments at the end of the survey. Thank you for your time and willingness to assist with this research project.

Please fill in the details below:

Position:

Police Office/Station/Agency:

Years of Efit services:

If you need assistance or have questions while taking this survey, please contact
Julie Gawryłowicz
j.gawrylowicz@abertay.ac.uk

Please read each question carefully and follow the instructions provided.

1) Which facial composite systems do you have experience with?
(Please tick all boxes that apply.)

Sketch artist	<input type="checkbox"/>	Efit-V	<input type="checkbox"/>	E-fit	<input type="checkbox"/>
Photofit	<input type="checkbox"/>	FACES	<input type="checkbox"/>	Evofit	<input type="checkbox"/>
Identikit	<input type="checkbox"/>	Profit	<input type="checkbox"/>	CDfit	<input type="checkbox"/>
Other	<input type="checkbox"/>				

(If 'other', please specify your answer below. If you need more room please use the additional space provided at the end of the questionnaire.)

2) Which facial composite system are you currently working with?
(Please tick all boxes that apply.)

Sketch artist	<input type="checkbox"/>	Efit -V	<input type="checkbox"/>	E-fit	<input type="checkbox"/>
Photofit	<input type="checkbox"/>	FACES	<input type="checkbox"/>	Evofit	<input type="checkbox"/>
Identikit	<input type="checkbox"/>	Profit	<input type="checkbox"/>	CDfit	<input type="checkbox"/>
Other	<input type="checkbox"/>				

(If 'other', please specify your answer below. If you need more room please use the additional space provided at the end of the questionnaire.)

- 3) How many years have you been working with the facial composite system indicated in Question 2?

(Please enter the appropriate number of years in the box.)

Year/s

(If less than a year, please enter the appropriate number of months in the box.)

Month/s

- 4) Did you receive training in the use of the facial composite system indicated in Question 2?

(Please tick the box that applies.)

Yes No (If you selected NO please go to Question 7.)

- 5) What kind of training did you receive?

(Please tick all boxes that apply.)

Training at the Scottish Police College Training from another officer in station or precinct
 National training centre in Durham Other

(If 'other', please specify your answer below. Please write your answer down in the space below. If you need more room please use the additional space provided at the end of the questionnaire.)

- 6) How many years ago was the training at the Scottish Police College?

(Please answer the following question only if you engaged in this training.)

Year/s

(If less than a year, please enter the appropriate number of months in the box.)

Month/s

How many years ago was the training at the National Training Centre in Durham?

(Please answer the following question only if you engaged in this training.)

Year/s

(If less than a year, please enter the appropriate number of months in the box.)

Month/s

- 9) Please estimate how many facial composites you personally have generated during the last 2 years.

(Please enter the appropriate number in the box.)

- 10) Please estimate how many facial composites your department have generated during the last 2 years.

(Please enter the appropriate number in the box.)

- 11) In general, how easy do you find it to select facial features (e.g. face shape, hairstyle, eyes and nose) recalled by a witness during the composite construction phase?

(Please tick box that applies.)

Very easy

Not applicable

Easy

Difficult

Very difficult

- 12) Based on your experience and knowledge, how much does the construction of a facial composite depend on the language ability of the witness?

(Please tick box that applies.)

Very much

Much

Little

Very little

- 13) Based on your experience and knowledge, how much does the construction of a facial composite depend on the memory ability of the witness?

(Please tick box that applies.)

Very much

Much

Little

Very little

- 14) In general, how **detailed** are verbal facial descriptions provided by witnesses?

(Please tick box that applies.)

Very detailed

Detailed

Moderately detailed

Not detailed at all

Based on your experience and knowledge please indicate the extent to which witnesses face the following difficulties when generating a facial composite.

15) In general, witnesses have difficulties picturing the perpetrator's face in their mind.
(Please tick box that applies.)

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

16) In general, witnesses have difficulties putting into words the description of the perpetrator's face.
(Please tick box that applies.)

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

17) In general, witnesses have difficulties understanding the instructions provided by the operator.
(Please tick box that applies.)

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

18) In general, witnesses have difficulties selecting individual features of the face during the feature selection phase.

(Please tick box that applies.)

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

19) In general, witnesses have difficulties in constructing fine-grain changes (e.g. wrinkles, stubble, etc.) of the face.

(Please tick box that applies.)

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

20) Have you ever generated a facial composite with the assistance of a witness with learning disabilities?

(Please tick box that applies.)

- Yes No

If you have experience with witnesses with learning disabilities, please base your answer to the following questions on your experience. If you do not have experience with people with learning disabilities please respond with your best judgment.

21) Are there any aspects of the facial composite system indicated in Question 2 which are particular **suitable** for people with learning disabilities?

(Please tick box that applies.)

- Yes
- No
- I do not know

If **YES**, please specify your answer.

(Please write your answer down below. If you need more room please use the additional space provided at the end of the questionnaire.)

24) **Please read each sentence and indicate to what extent witnesses with learning disabilities face the following difficulties when generating a facial composite.**

a) In general, witnesses have difficulties picturing the perpetrator's face in their mind.
(Please tick box that applies.)

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

b) In general, witnesses have difficulties putting into words the description of the perpetrator's face.
(Please tick box that applies.)

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

c) In general, witnesses have difficulties understanding the instructions provided by the operator.
(Please tick box that applies.)

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree



UNIVERSITY
of
ABERTAY DUNDEE



Appendix 2

Male facial stimuli used during the description task in Study 2



Appendix 3

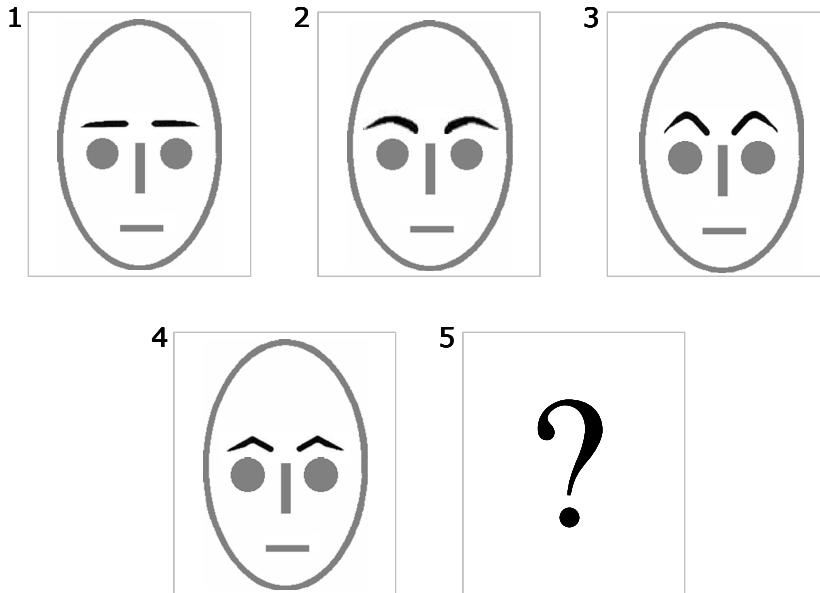
Facial stimuli used during the E-FIT composite construction in Study 3



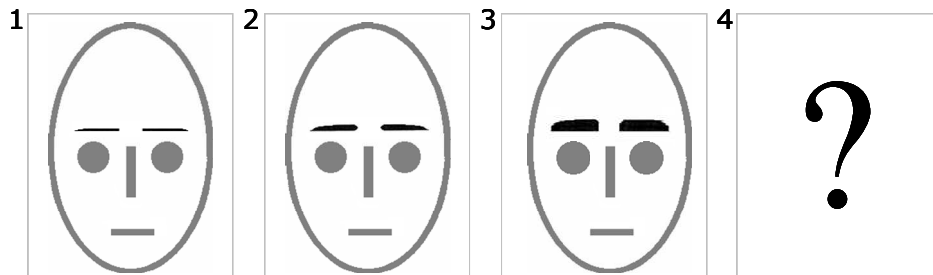
Appendix 4

The complete set of visual prompts used during the visual prompt task in Study 4

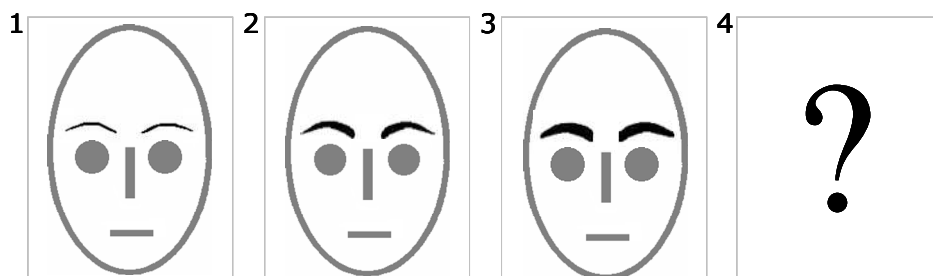
1.1 Eyebrow Shape



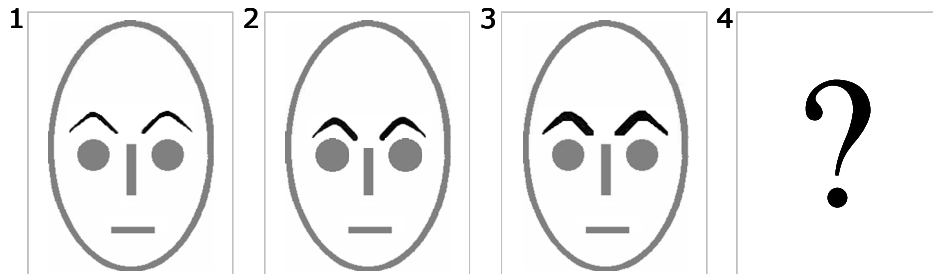
1.2 Eyebrow Size (shape 1)



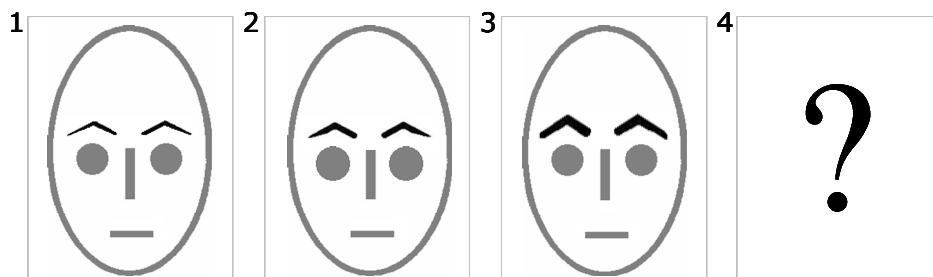
1.2 Eyebrow Size (shape 2)



1.2 Eyebrow Size (shape 3)



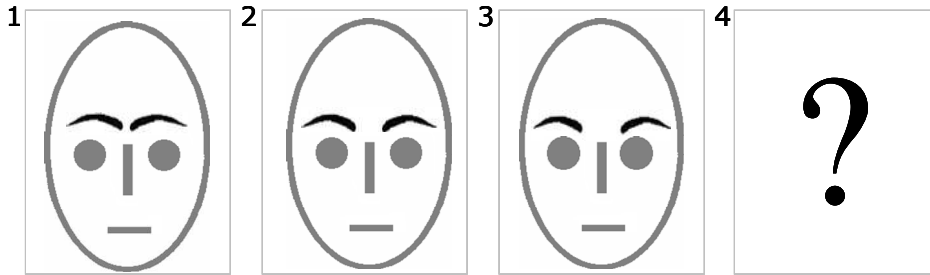
1.2 Eyebrow Size (shape 4)



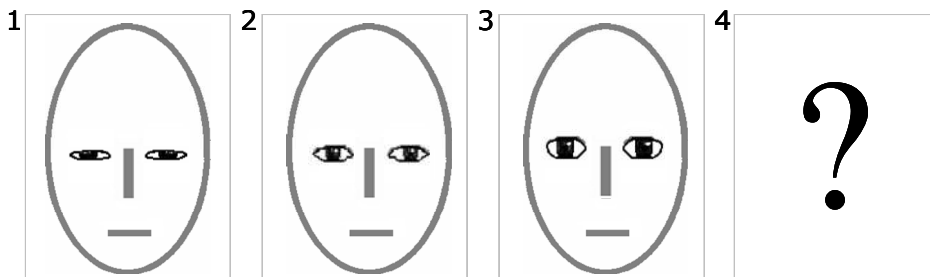
1.3 Eyebrow Colour



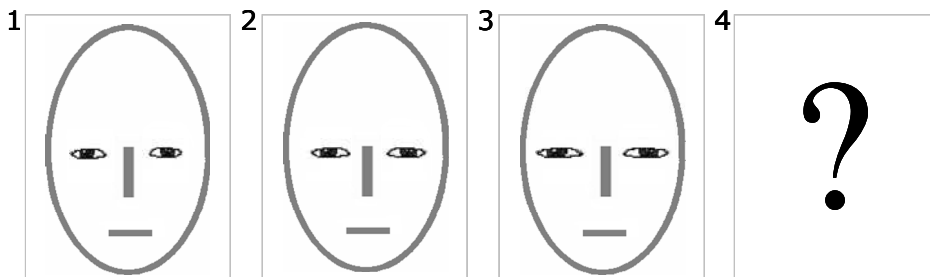
1.4 Eyebrow Spacing



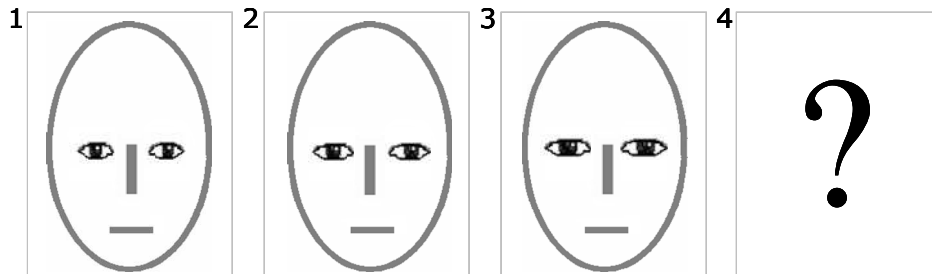
2.1 Eye Shape



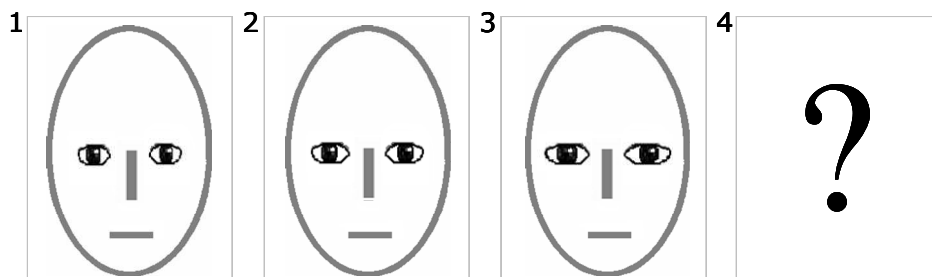
2.2 Eye Size (shape 1)



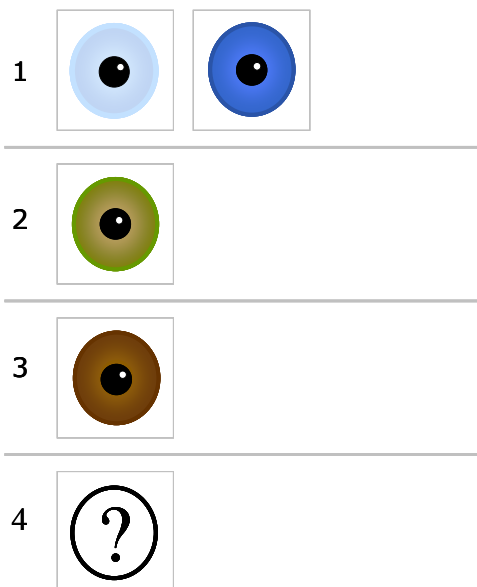
2.2 Eye Size (shape 2)



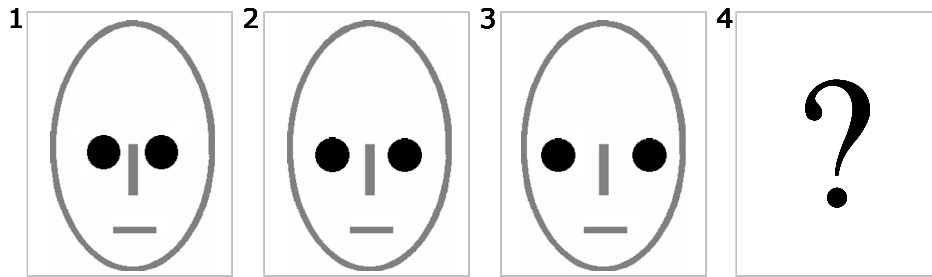
2.2 Eye Size (shape 3)



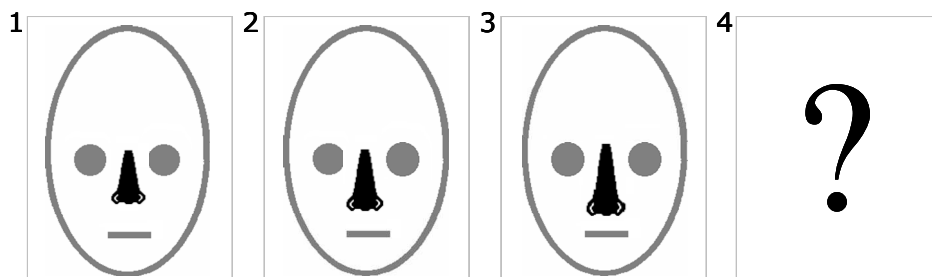
2.3 Eye Colour



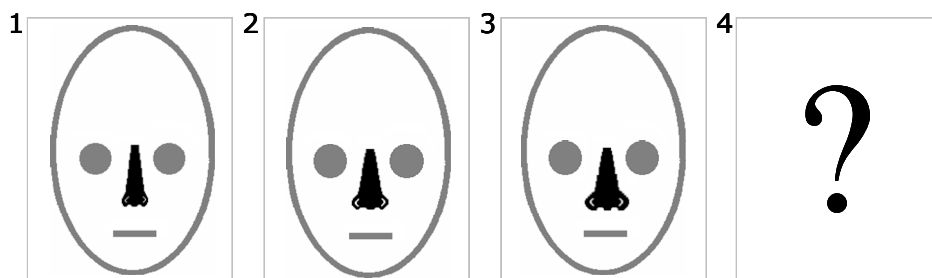
2.4 Eye Spacing



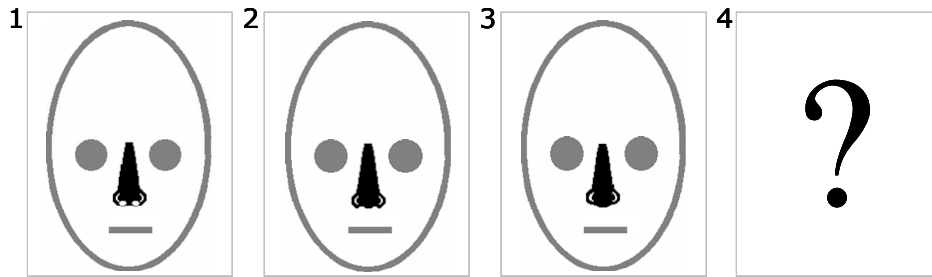
3.1 Nose Length



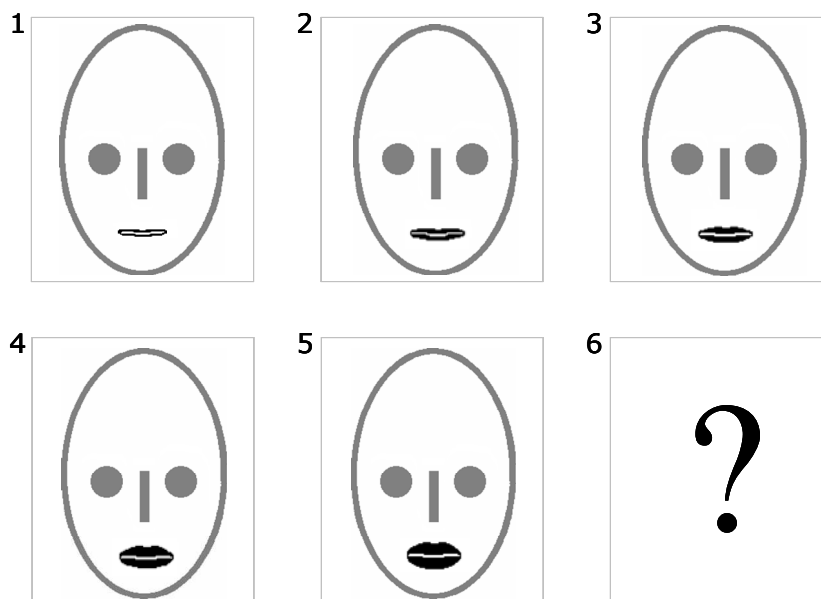
3.2 Nose Breadth



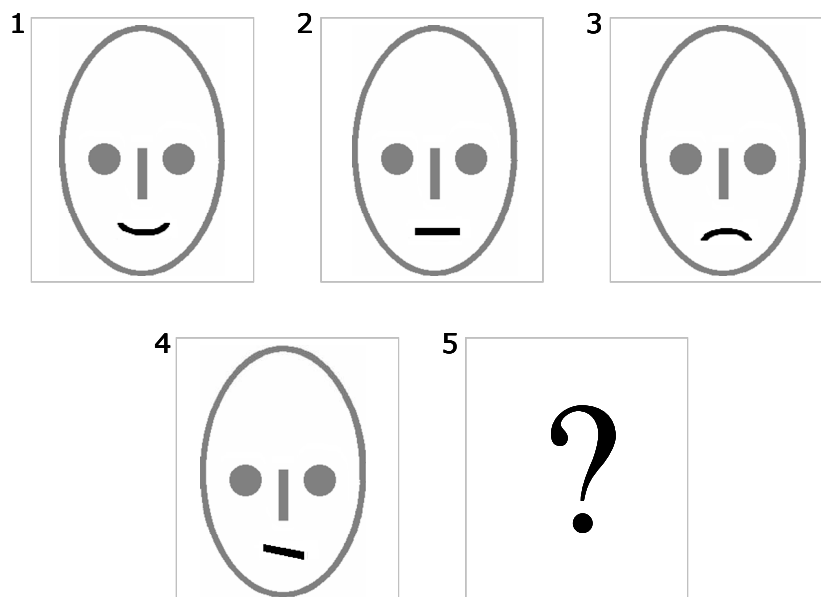
3.3 Nose Tip



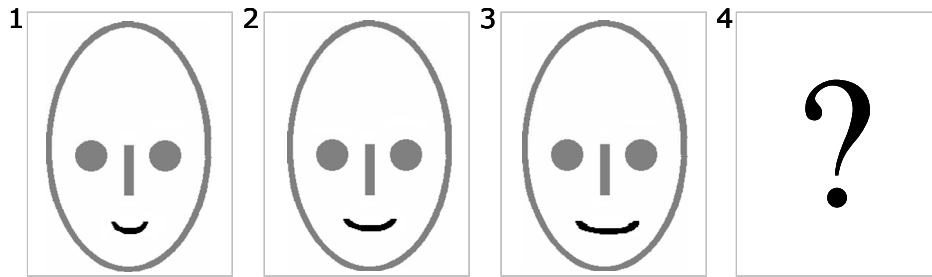
4.1 Lips



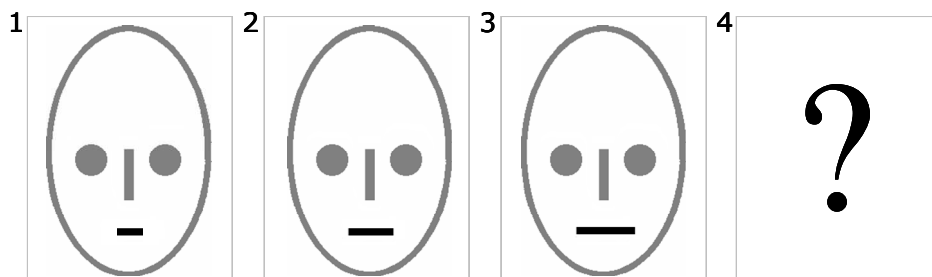
4.2 Mouth Shape



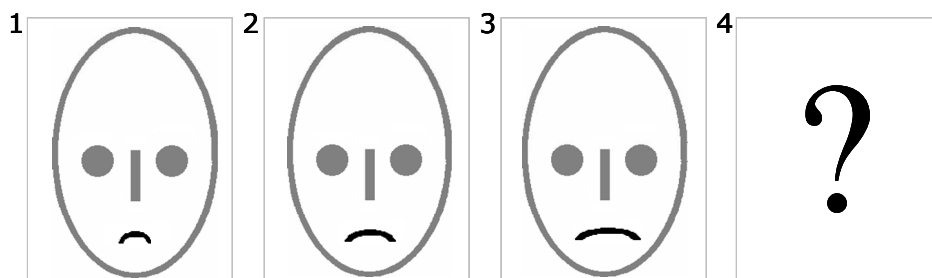
4.3 Mouth Size (shape 1)



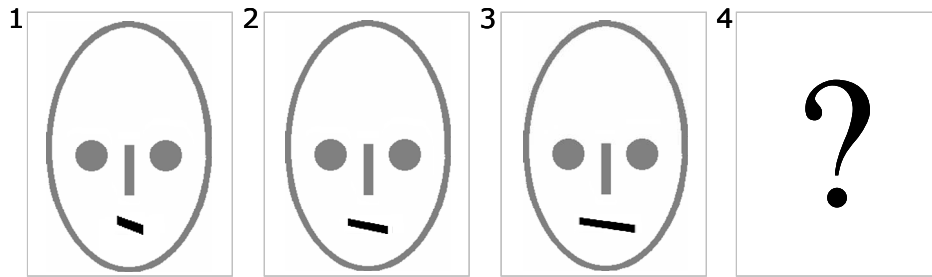
4.3 Mouth Size (shape 2)



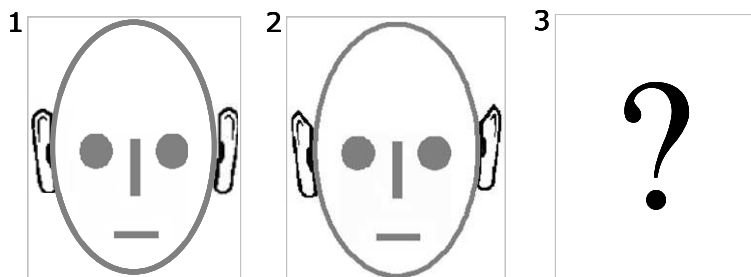
4.3 Mouth Size (shape 3)



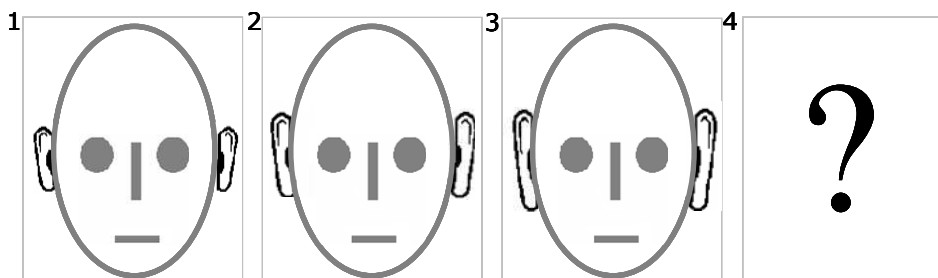
4.3 Mouth Size (shape 4)



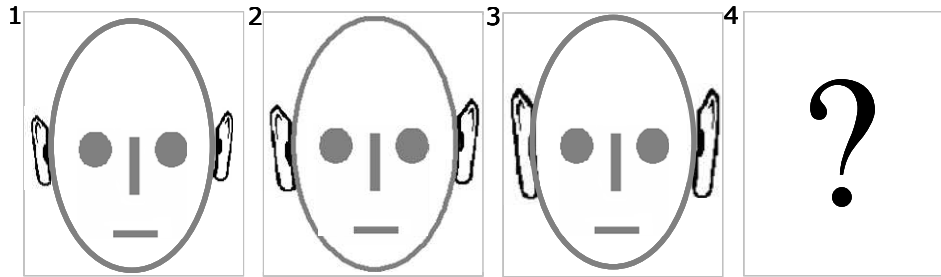
5.1 Ear Shape



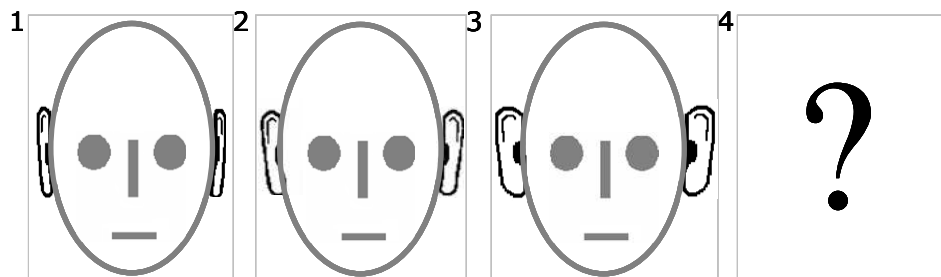
5.2 Ear Size (shape 1)



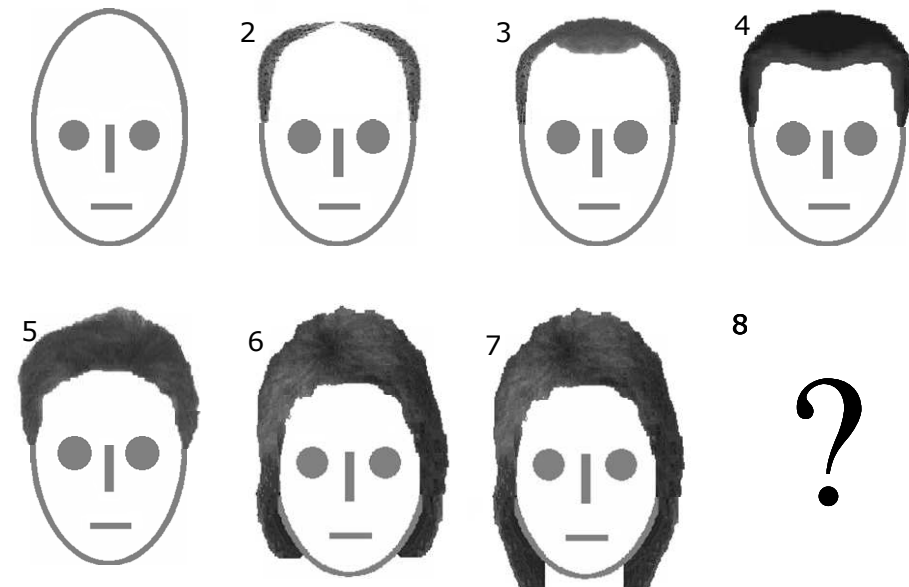
5.2 Ear Size (shape 2)



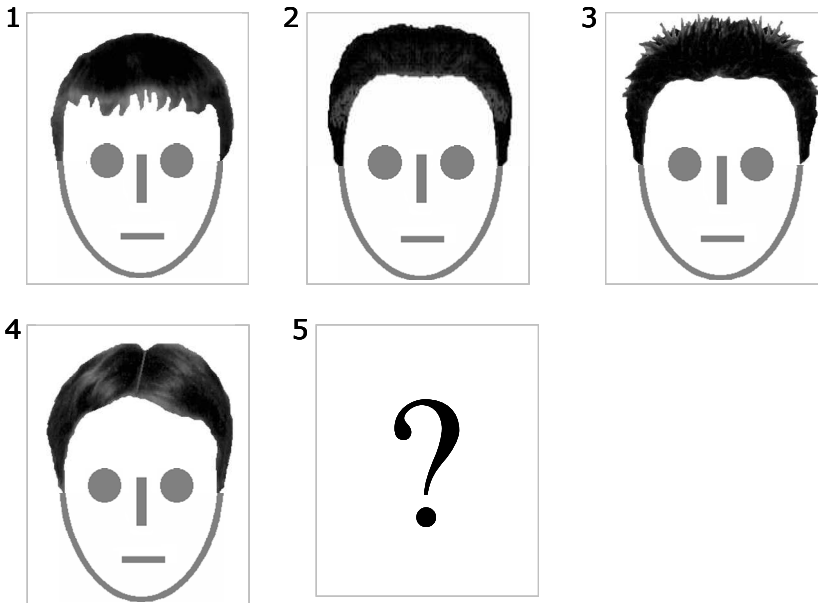
5.3 Ear Setting



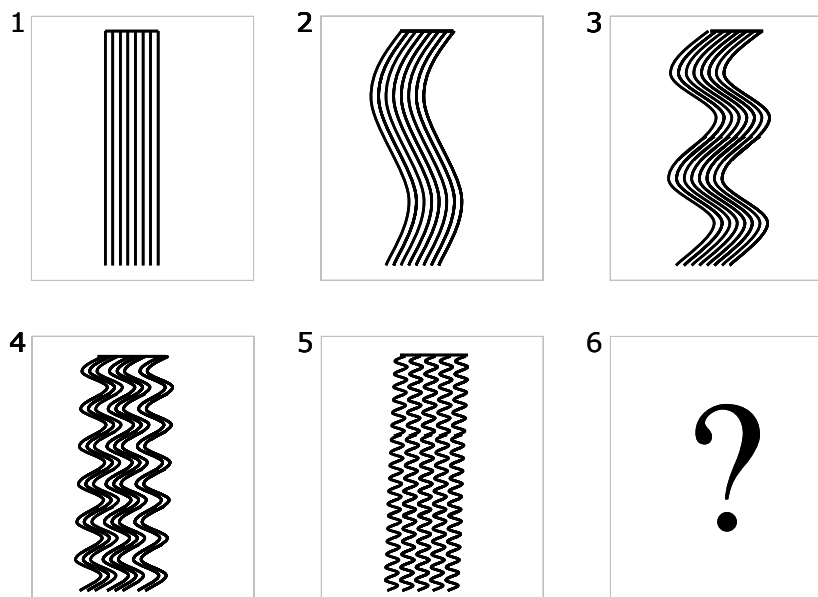
6.1 Hair Thickness / Length



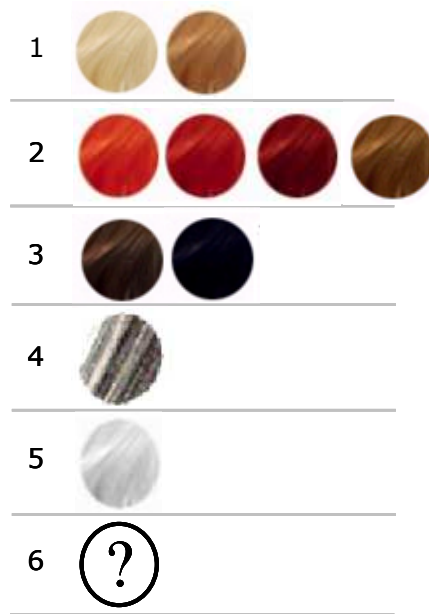
6.2 Hair Brushing / Style



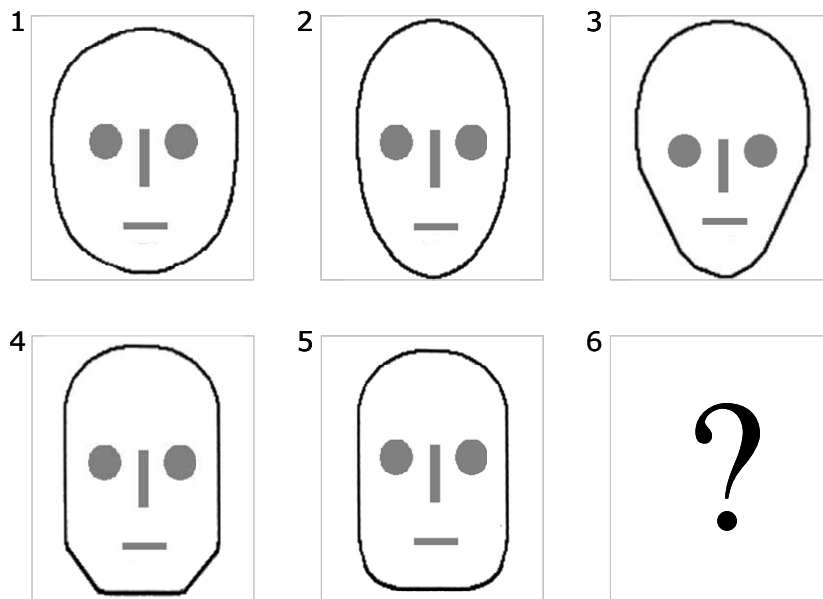
6.3 Hair Type



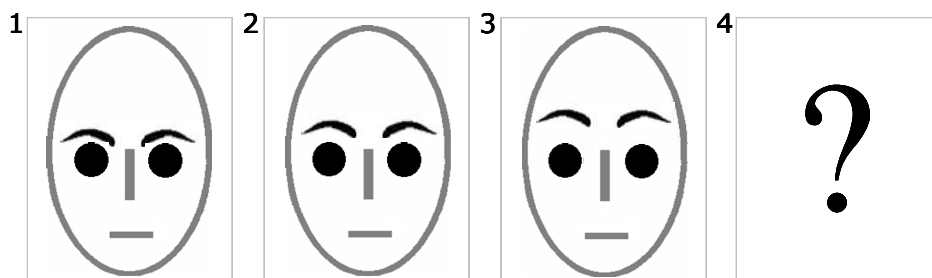
6.4 Hair Colour



7.1 Face Shape



8.1 Eye & Eyebrow Spacing



9.1 Nose, Mouth & Chin Spacing

