TRYING TO IMPROVE CHILD AND YOUNG ADULT WITNESSES' PERFORMANCE

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Kristjan Kask, MSc (University of Tartu, Estonia)
School of Psychology
University of Leicester

January 2008

UMI Number: U521519

All rights reserved

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

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

UMI®
Dissertation Publishing

UMI U521519
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest®

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

Dedicated to my parents who always kept believing in me whatever and whenever I did.

ACKNOWLEDGEMENTS

First of all I'd like to thank my supervisor Ray Bull who was there all the time when needed and always read and discussed my studies and proposals very eagerly. I know I'm going to miss our insightful conversations a lot. I'd also like to thank my second supervisor Graham Davies who gave useful hints how to organize the thesis and what to do to get the studies published.

Many thanks for Raphael Gillett for supporting me on the roads of statistics; Tony Andrews on provided excellent services concerning computers and IT; and Tim Grant for useful hints at the half-way 'mini-viva' stage of my PhD.

Concerning the completion of different studies, I am grateful to all the authors who responded my emails concerning Study 1. Many thanks to Piia Enno and especially to Pille Alaver from Estonian Police Force for their assistance during the data collection of Study 2; also to the young adults, their teachers (especially Nick Oliver) and staff of Queen Elizabeth I & Wiggeston College in Leicester, UK, and also to Indrek Heinla, Anneli Kompus, Clara Novo, Matthew Aylott, and Alfonso Martinez for their help gathering the data for Study 3.

Concerning Studies 4 and 5 I'd like to thank Siegfried L. Sporer (University of Giessen) and to Juergen Gehrke (Nottingham Trent University) for their help during the course of this research; and to children and young adults, their head teachers and staff of high schools in Antsla, Põlva, Otepää, and Tallinn in Estonia.

And last but not least, many thanks to my parents, Liisi, and all of my friends for keeping me going and for their support in better and in worse. Also I'd like to thank Archimedes Foundation from Estonia who kindly sponsored and supported my studies and without whom I would not have been able to study for this PhD degree in Leicester.

THESIS ABSTRACT

In the past decades a large amount of research has been conducted on child witnesses. As the number of children who testimony in court increases, the need to obtain accurate information has not played a more central role in forensic research than at the present time. Considerable research has previously been undertaken to increase the quality and quantity of information in children's accounts, there is still a lack of knowledge and methods to improve children's performance in forensic context.

Nevertheless, the present PhD thesis aimed to examine possible ways of trying to improve child and young adult witnesses' testimony. The author was particularly interested in questioning methods with children, young adults' person descriptions, and face recognition and identification ability of both children and young adults (including the ability to recognize different race faces). Chapter 1 reviews the relevant literature on issues with child witnesses and the aims of the present thesis (submitted to Trames).

Chapter 2 presents a meta-analysis of identification studies involving children. Chapter 3 (submitted to Child Abuse and Neglect) examines Estonian investigators' questioning styles with child witnesses focusing on the length and type of details in children's answers to questions. Chapter 4 (published in Psychiatry, Psychology and Law, 2006) examines the effects of using a person in young adult witnesses' visual field as a comparison (or ´standard`) to assist their recall of a previously seen, different person. Chapters 5 and 6 (in press, Psychology, Crime, and Law) present the comparison of Estonian children's and young adults' face recognition ability of different race faces, as there appear to have been published no studies of the cross-racial effect in less ´Westernised` societies, such as in Estonia. Also new sequential target presentation methods were used.

All results are discussed with reference to previous findings in Chapter 7. In that chapter methodological shortcomings and statistical problems are reviewed with thoughts on possible future research and an evaluation of present knowledge on child witness testimony.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv
THESIS ABSTRACT v
TABLE OF CONTENTS vi
LIST OF TABLES xi
LIST OF FIGURES xiii
1. CHAPTER 1: ISSUES IN CHILDREN'S TESTIMONIES 1
1.1. INTRODUCTION 1
1.2. THEORETICAL ISSUES IN DESCRIPTIONS 1
1.3. CONTENT OF PERSON DESCRIPTIONS 4
1.4. CHARACTERISTICS OF WITNESS 6
1.4.1. Height, weight, and age 6
1.4.2. Gender of witness 6
1.5. DESCRIPTION-IDENTIFICATION RELATIONSHIP 7
1.6. FACE RECOGNITION AND FEATURE CHECKLISTS 7
1.7. VERBAL OVERSHADOWINGS AND VERBALIZATION 8
1.8. EMPIRICAL RESEARCH WITH CHILDREN 10
1.8.1. Height, weight, and age 12
1.8.2. Gender of witness 12
1.9. CHILDREN'S FACE RECOGNITION 13
1.9.1. Own-age bias 14
1.10. QUESTIONING METHODS WITH CHILDREN 15
1.11. CONCLUSION 16
2. CHAPTER 2: A META-ANALYSIS OF IDENTIFICATION STUDIES INVOLVING CHILDREN. 18
2.1. ABSTRACT 18
2.2. INTRODUCTION 19
2.2.1. Factors moderating identification accuracy 20
2.2.2. Target-present and target-absent lineups 21
2.2.3. Method of lineup presentation 21
2.2.4. The mode of target presentation 23
2.2.5. Delay 24
2.2.6. Number of members in the lineup 24
2.2.7. Aims of the meta-analysis. 24

2.3. METHOD 25
2.3.1. Sample 25
2.3.2. Study characteristics 26
2.3.3. Dependent variable 27
2.3.4. Statistics 27
2.4. RESULTS 28
2.4.1. Target present lineups 28
2.4.2. Target absent lineups 34
2.4.3. Comparison of correct identifications and false alarms 39
2.5. DISCUSSION 39
2.5.1. Age effects 40
2.5.2. The number of members in a lineup 41
2.5.3. The effect of delay 41
2.5.4. The type of lineup 41
2.5.5. The mode of target presentation 42
2.5.6. Comparison of correct identification and false alarm rates 42
2.6. CONCLUSION 42
3. CHAPTER 3: THE AMOUNT OF JUDICIALLY SIGNIFICANT DETAILS IN ESTONIAN INVESTIGATORS' QUESTIONING STYLES WITH CHILD WITNESSES 44
3.1. ABSTRACT 44
3.2. INTRODUCTION 45
3.2.1. Quality of interviews 45
3.2.2. Verbal ability 47
3.2.3. Children's responses 48
3.2.4. Present study 49
3.3. METHOD 49
3.3.1. Participants 49
3.3.2. Coding 50
3.3.3. Details reported by the child 52
3.4. RESULTS 53
3.4.1. The proportion of questions in interviews 53
3.4.2. Children's responses to the questions 55
3.4.3. Type of information reported by the children 56
3.4.4. An international comparison of the results 57

3.5. DISCUSSION 58
3.5.1. Type of information reported by the children 60
3.5.2. An international comparison of the results 61
3.5.3. Suggestions for the future 61
3.6. CONCLUSION 64
4. CHAPTER 4: TRYING TO IMPROVE YOUNG ADULTS' PERSON DESCRIPTIONS 65
4.1. ABSTRACT 65
4.2. INTRODUCTION 66
4.2.1. Why are people poor at describing other people? 66
4.2.2. Attempts to improve (children's) person descriptions 67
4.2.3. Gender differences in person descriptions 69
4.2.4. Cross-ethnic effects 69
4.2.5. Own-anchor effect 70
4.2.6. Present study 71
4.3. METHOD 71
4.3.1. Participants 71
4.3.2. Design 71
4.3.3. Procedure 71
4.3.4. Coding 73
4.4. RESULTS 73
4.4.1. Descriptions 74
4.4.2. Answers to the questions 74
4.4.3. The effect of type of standard 74
4.4.4. Effect of use of a standard regarding 'easy' and 'difficult' characteristics 75
4.4.5. Own-anchor effect 79
4.4.6. Gender differences on own-anchor effects 79
4.5. DISCUSSION 79
4.5.1. Young adults' face descriptions 79
4.5.2. Using the standard in answers to the questions. 80
4.5.3. Own-anchor effect 80
4.6. CONCLUSION 81
5. CHAPTER 5: ESTONIAN CHILDREN'S AND YOUNG ADULTS' 82
5.1. ABSTRACT 82

5.2. INTRODUCTION 83
5.2.1. Different mechanisms in cross-race face recognition 84
5.2.2. Social and cognitive mechanisms in the cross-race effect 84
5.2.3. Children's face recognition ability and cross-race effect 88
5.2.4. Face description ability 90
5.2.5. Present study 91
5.3. METHOD 91
5.3.1. Participants 91
5.3.2. Design 92
5.3.3. Procedure 92
5.3.4. Test faces 93
5.3.5. Coding 93
5.4. RESULTS 93
5.4.1. Effects of different races. 94
5.4.2. Age differences in the recognition of different race faces 95
5.4.3. Choosing target or most similar foil in the test set. 96
5.4.4. Age differences in choosing target or most similar foil in test set. 96
5.4.5. Young adults' face descriptions 97
5.5. DISCUSSION 97
5.5.1. Differences between face recognition of different races. 97
5.5.2. Young adults' face descriptions 101
5.6. CONCLUSION 101
6. CHAPTER 6: A COMPARISON OF THREE SEQUENTIAL PRESENTATION METHODS REGARDING OTHER-RACE FACE RECOGNITION 103
6.1. ABSTRACT 103
6.2. INTRODUCTION 104
6.2.1. Sequential lineups 104
6.2.2. Multiple choice presentation methods 105
6.2.3. Cross-race effect in multiple presentations 106
6.2.4. Gender differences in cross-race face recognition. 107
6.2.5. Present study 107
6.3. METHOD 108
6.3.1. Participants 108
6.3.2. Design 108
6.3.3. Procedure 109

6.3.4. Test faces 110
6.4. RESULTS 110
6.4.1. Effects of different races. 111
6.4.2. Gender effects in different race face recognition 112
6.4.3. Effect of different presentation methods on face recognition 112
6.4.4. Effect of gender on different presentation methods 113
6.5. DISCUSSION 114
6.5.1. Effects of different races. 115
6.5.2. Effect of different presentation methods on face recognition 115
6.5.3. Gender effects 117
6.6. CONCLUSION 117
7. CHAPTER 7: OVERVIEW OF TRYING TO IMPROVE YOUNG PERSONS' WITNESS PERFORMANCE. 119
7.1. INTRODUCTION 119
7.2. SUMMARY OF EMPIRICAL CHAPTERS 119
7.2.1. Study 1. 119
7.2.2. Study 2. 121
7.2.3. Study 3. 122
7.2.4. Study 4. 123
7.2.5. Study 5. 125
7.3. METHODOLOGICAL AND STASTISTICAL ISSUES 127
7.4. FUTURE RESEARCH 131
7.5. CONCLUSION 134
8. REFERENCES 136
9. APPENDICES 170
9.1. Appendix 1: Questions asked using or not using the standard in Study 3 170
9.2. Appendix 2: Drawings representing (from top to bottom) faces, hairstyles, mouths, noses, and figures in the using ´standard` mode in Study 3. 172
9.3. Appendix 3: Targets and their target replacements (below each) in Study 4 173
9.4. Appendix 4: Stimulus sheets in Study 5 175
9.5. Appendix 5: Face recognition sets in Study 5. 176
9.6. Appendix 6: Targets and their target replacements (below each) in Study 5 178
9.7. Appendix 7: Stimulus sheets in Study 5 179
9.8. Appendix 8: Face recognition sets in Study 5. 180

LIST OF TABLES

Chapter 2.

Table 2.1. Target-Present Hypothesis Tests. 30

Table 2.2. Target-Absent Hypothesis Tests. 36

Chapter 3.

Table 3.1. The Proportion of Questions in Interviews. 53

Table 3.2. Quotations from the Interviews. 54

Table 3.3. The Proportion of Questions (in percentages) in Different Quarters of Interviews. 55

Table 3.4. The Number and the Proportion of Information Units in Response to Different Categories. 56

Table 3.5. The Number of Contextual and Judicially Significant Details Reported in Response to Different Categories of Interviewer Utterances. 57

Table 3.6. The Frequency of Different Interviewer Utterance Types in Percentages: An International Comparison. 58

Chapter 5.

Table 5.1. Percentages and Frequencies of Correct Identifications, False Identifications, and Correct Rejections 94

Table 5.2. The Percentages and Frequencies of Correct Identifications, False Identifications, and Correct Rejections in Children and Adults. 95

Table 5.3. Percentages and Frequencies of Choosing Target or Foil First. 96

Table 5.4. Percentages and Frequencies of Adults and Children Choosing Target or Foil First. 97

Chapter 6.

Table 6.1. Percentages and Frequencies of Correct Identifications, False Identifications, and Correct Rejections of Own- and Other-Race Faces. 111

Table 6.2. Gender Effects and Frequencies in Correct Identifications, False Identifications, and Correct Rejections. 112

Table 6.3. Percentages and Frequencies in Correct Identifications, False Identifications, and Correct Rejections Across Different Presentation Methods 113

Table 6.4. Gender Effects and Frequencies in Correct Identifications, False Identifications, and Correct Rejections across Different Presentation Methods. 114

LIST OF FIGURES

Chapter 1.

Figure 1.1. Verbal and visual translation process between witness and investigator (after Sporer, 2001b). 3

Chapter 4.

Figure 4.1. The effect of type of standard and no standard in answering questions. 75

Figure 4.2. The effect of type of standard and no standard for ´difficult` Characteristics. 76

Figure 4.3. The effect of type of standard and no standard for ´easy` characteristics. 76

Figure 4.4. The effect of type of standard and no standard in medium frequency characteristics. 77

Figure 4.5. The effect of type of standard and no standard in high frequency characteristics. 78

Figure 4.6. The effect of type of standard and no standard in low frequency characteristics. 78

Chapter 5.

Figure 5.1. In-group/out-group model of face processing (Sporer, 2001a). Image from Sporer (2007). 87

1. CHAPTER 1: ISSUES IN CHILDREN'S TESTIMONIES

1.1. INTRODUCTION

In the past 25 years, more and more research has been conducted on child witnesses. Young children are increasingly being called to testify in criminal cases, particularly in sexual abuse cases (Schepard, 2004). However, surveys of adults' beliefs about child eyewitnesses have suggested that the general population has a pessimistic view of children's eyewitness capabilities (Yarmey & Jones, 1983; Ross, Dunning, Toglia, & Ceci, 1990; Nikonova & Ogloff, 2005) and that adult witnesses are rated to be more credible than child witnesses (Pozzulo, Lemieux, Wells, & McCuaig, 2006).

This chapter reviews research concerning child witnesses that is relevant to the (new) studies to be reported in this thesis. The first part of the chapter focuses on adults because with regard to several factors that may affect child witnesses, only studies with adults have so far been conducted to date. First, an overview is given of theoretical issues regarding person descriptions such as verbal and visual processing, and characteristics influencing these processes, followed by a review of archival and empirical studies of person descriptions by adults and characteristics of witnesses. After that, issues regarding gender differences, the description-identification relationship, face recognition, and the concepts of verbal overshadowing and verbalization are covered.

In the second part of the chapter, studies on children's person descriptions are reviewed, followed by an overview of studies regarding children's face recognition ability. Finally, the outcomes of different questioning methods are reviewed. In this chapter 'children' is a term used to cover all those below 14 years of age.

1.2. THEORETICAL ISSUES IN DESCRIPTIONS

Eyewitnesses play a crucial role in bringing perpetrators of crime to justice. The police rely very much on witness testimony, especially at the beginning of their investigations (Fisher & Geiselman, 1992). Collecting person descriptions from witnesses is also one of the regular information-collecting tools (Meissner, Sporer, & Schooler, 2006).

Person descriptions are usually verbal reproductions of (visually) perceived stimuli (Sporer, 1996). Such descriptions generally contain references to physical appearance (mainly face, height and weight), stature and clothing. Descriptions regarding the face

can be crucial for identifying a person. However, as Sporer (1996) states, it is hard to translate a rich visual impression into a detailed verbal description (whereas the identification of a person at an identification parade, or from a photo spread is an act of visual recognition which is considered to be easier). Our vocabulary for expressing the physical aspects of faces is rather limited, when compared to the large number of adjectives available for describing character traits (Shepherd & Ellis, 1996). It has been found that respondents tend toward character attributes in their descriptions of a once-seen person, even when they have been explicitly instructed to provide only physical descriptions (Sporer, 1996).

One explanation of why it is hard to recall physical aspects of faces rather than character traits, was provided by Paivio (1969, 1971) who suggested that verbal information is processed differently from imaginal (i.e. visual) information. However, he also contended that we use both imaginal and verbal codes to some extent for representing either sort of information (dual coding). According to Paivio (1969), mental images are analogue codes (a form of knowledge representation that preserves the main perceptual features of whatever is being represented) for the physical stimuli, but our mental representations for words are represented in a symbolic code (a form of knowledge representation that stands for something and does not perceptually resemble whatever is being represented).

Breznitz (2002), among others, has claimed that the auditory (e.g. verbal) and visual systems process information differently. Discrimination and identification are achieved faster through the visual than the auditory route because the visual system processes information holistically whereas the auditory system processes information sequentially. As the visual route is faster and holistic, it could be applied also to other processes, like face or person recognition. This partly explains why we have difficulties in coding visually retrieved information into a verbal mode.

The visual and verbal encoding and decoding processes involved in an exchange between the witness and the investigator could be as follows (Sporer, 1996). For effective communication, the demarcation between the visual and verbal modes has to be transgressed twice (see Figure 1.1). Firstly, the witness has to transform a visual impression of the target into a verbal person description. One of the problems here is

that faces are best encoded holistically (Wegner & Ingvalson, 2002), whereas a verbal description uses more a piecemeal approach involving the 'labeling' of the individual features (where the object of description is broken down into singular details) (Sporer, 1992a). The police usually require featural rather than holistic descriptors.

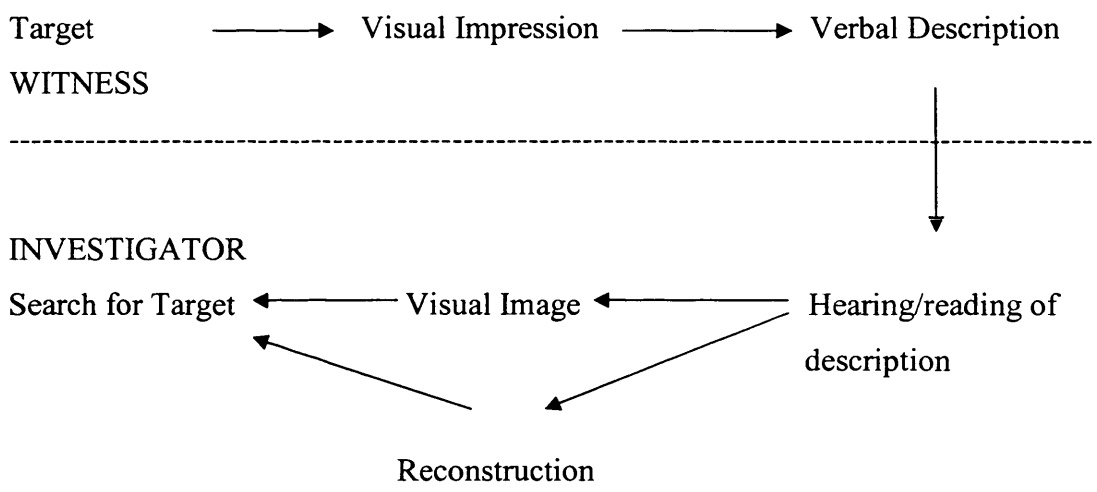


Figure 1.1. Verbal and visual translation process between witness and investigator (after Sporer, 1996).

Secondly, the description is conveyed verbally to the investigator who usually has to re-transform this description into a visual representation of the described person. However, in some cases investigator just has to pass the person description to the officers in the field and then they have to construct a visual image of the culprit based on the description. (The re-transforming process might not happen as described above when the investigator is searching police databases for culprits in similar crimes and then tries to match verbal description to the images of faces in the database.)

One problem with such transformation / re-transformation is that a witness may describe a person as having a 'hooked nose', but for somebody else to hear/read this description there is an almost infinite number of curvatures imaginable that would still be compatible with the description 'hooked nose', but no longer to the nose originally perceived by the witness. Or if a witness mentioned character traits such as 'good-looking' or 'he looked like a bank manager' - investigators could 'perceive' this also differently (using rather a piecemeal approach). This process can be applied to most featural descriptors and could be the source of mistakes when constructing an image of the described person. Finally, despite the difficulties translating a visual image into a

verbal description, the police still tend to use person descriptions in their every-day work as one of the primary information-collecting tools (Meissner, Sporer, & Schooler, 2006) or in identification parades which are based on the witness' description of the culprit (McQuiston-Surrett, Malpass, & Tredoux, 2006).

1.3. CONTENT OF PERSON DESCRIPTIONS

Empirical studies regarding the content of person descriptions begun over 30 years ago with archival study by Kuehn (1974) who analyzed person descriptions provided by victims in 100 police files and found that mostly gender, age, height, build, race, weight, complexion, and hair colour were mentioned (such features were mentioned by more than 70% of victims, the mean being 7.2 descriptors). These features could be better remembered because they may play more crucial role later in person recognition. For example, Shepherd (1981) noted that hair, face outline, eyes and mouth are important for perceiving and remembering faces. Seitz (2002) found that eyes and mouth were more important features than nose in accurate face recognition.

Yuille and Cutshall (1986) examined police files regarding a single shooting incident and they found 392 action, 180 person description, and 78 object description details (of which 82%, 76%, and 89% turned out to be correct). Most of the errors were for person descriptions regarding height, weight, and age (which have been found hard to estimate for both adults and children). There was a lack of memory loss over time, perhaps due to rehearsal, although some specific aspects of the event were forgotten (e.g. some colours, particularly referring to clothing, were not well remembered over time).

Sporer (1992b) in a content analysis of crime files regarding 100 witnesses found that the average number of items provided in the person descriptions was 9.71 and that 22.4% of the descriptive details referred to general features such as race, age, height, stature, and movements. Another 31% of the descriptors were about clothes, 29.6% about the face (mainly hair and beard), 5% mentioned personality inference, and 12% of ´other` features (e.g. jewellery, dialect, disguise, smell).

Van Koppen and Lochun (1997) found that actual forensic witnesses mentioned an average of eight descriptors which referred to more general features such as sex, race, and build, than particular to facial characteristics (which have been found harder to

describe probably due to lack of vocabulary). The most frequently mentioned characteristics were the gender and height of the perpetrator. Reports of these particular descriptors were completely or partly accurate for more than 80% of the witnesses and witnesses were fairly accurate in their descriptions of age, build, height, and hair. Descriptions contained more permanent (e.g. gender, race) than temporary characteristics (e.g. clothing, disguises) and overall, of the descriptions, 59% were correct, 17% partially correct, and 24% incorrect. Witnesses were more incorrect regarding characteristics such as inner face features, dialect, and type of hair. Explanations of this last finding could be that these characteristics are harder to describe due either to a lack of vocabulary or to difficulties in decision-making (i.e. deciding what type of dialect it was). Also, these characteristics can be subjective (i.e. vary across describers) because of variability in how they verbalize them.

Comparing real crimes and staged crimes, Lindsay, Martin, and Weber (1994) found that witnesses viewing *staged crimes* mostly reported clothing (99%), hair color (90%) and weight (86%), and the most frequently reported facial feature was the eyes (43%). However, witnesses of *real crimes* mostly reported gender (96%), hair color (38%) and clothing (60%), while facial features were reported by less than 10% of the sample. Also, staged crime witnesses provided more complete descriptions than did real witnesses (7.35 vs. 3.94). This difference could be due to various factors negatively influencing encoding in real crimes.

Some other characteristics also affect the quality of person descriptions. *Time delay* between the event and accurately communicating a visual impression to an investigator can influence the quality of testimony. Van Koppen and Lochun (1997) found that witnesses provided fewer person descriptors following longer retention intervals (and that better illumination and shorter distances between the witness and perpetrator were associated with better person descriptions). However, Yuille and Cutshall (1986), found high levels of recall from witnesses of a real crime two years after the incident. Nevertheless, laboratory research has consistently shown significant effects of delay on the accuracy of person descriptions.

1.4. CHARACTERISTICS OF WITNESS

1.4.1. Height, weight, and age

As stated above, many witness descriptions contain characteristics such as height, weight, and age of the culprit. However, tall and heavy targets might be underestimated, while short and light targets might be overestimated. This would reflect a general regression toward the mean. Flin and Shepherd (1986) found that observers' own height and weight tended to effect their estimations of the perpetrators' height and weight, but this 'own-anchor' effect was found only for male observers describing a male target (see Study 3 in Chapter 4 for a fuller overview). Thus, witnesses' own characteristics might influence their descriptions (Study 3 examines this possibility). Also, the estimations about physical appearance characteristics of perpetrators might also be influenced by witness knowledge about population norms (Meissner, Sporer, & Schooler, 2006).

Manis and Paskewitz (1984) have proposed a judgment model which suggests that prior experience with some members of a given category affects the assessment of other members of the same category in two ways: (i) by providing a basis for comparison and (ii) by leading the judge to expect that new exemplars will resemble those previously encountered. It is generally accepted that estimates of height and weight are relatively inaccurate and thus might be influenced by personal determinants (see Clifford & Bull, 1978; Janssen & Horowski, 1980). Biernat, Manis, and Nelson (1991) found an own-gender effect when people were estimating persons' height and weight. They found that when an average college student was the standard (with whom to evaluate the person's height and weight estimates in photographs), then male targets were rated in comparison with other men and female targets with other women.

1.4.2. Gender of witness

It has been found that gender of witness can affect descriptions. Women have been found superior to men in recall for targets' weight, hair colour and length (Yarmey, 1993), these being attributes that women may have been more likely to attend to at encoding. Powers, Andriks, and Loftus (1979) found that women were more accurate than men on questions dealing with women's clothing or actions, whereas men were

more accurate on questions concerning men's appearance and surroundings. Sporer (1996) reported that on average male witnesses provided longer descriptions than females (but this does not mean that females convey information that is less accurate).

1.5. DESCRIPTION-IDENTIFICATION RELATIONSHIP

It seems intuitive that witnesses who are better at describing a perpetrator should also be better at identifying him/her (Meissner et al., 2006). Despite the belief that a strong relationship should exist between face description quality and identification accuracy, research reveals that such a relationship is at best very weak. Numerous studies demonstrate that the ability to describe people does not correlate highly with the ability to recognize people (e.g., Sporer, 1989, 1996; Wells, 1985; Meissner, Brigham, & Kelley, 2001). There is only one study which has found that opposite, namely Sporer (1992b) has reported significant positive relations between the number of descriptors each participant provided of a face and later identification accuracy of that face (in a staged event study).

Meissner et al. (2006) note that the absence of a relationship between the accuracy of person description and identification may provide an important link in our understanding of these two tasks, namely that cognitive processes concerning person descriptions are very different from those involved in the identification of a face. They suggested that person descriptions may encourage focusing upon verbalizable features of the face that are not always useful for perceptually individuating a given face from among similar distractors. Farah, Wilson, Drain, and Tanaka (1998) claim that recognition of faces involves a configural process in which features combine to create a non-verbal perceptual set that is stored and later accessed for pattern recognition. In short, face description could distort the veracity of the memory trace and interfere with subsequent identification.

1.6. FACE RECOGNITION AND FEATURE CHECKLISTS

Ellis (1990) found that less than 5% of descriptors that the participants spontaneously provided referred to inner features of the face (which are considered to be among the most important factors in identifying another person, see also Seitz, 2002). Fisher and Cox (1975) found that the majority of the adults referred to the upper half of the face.

Bruce and Young (1986) developed a model of face recognition which uses information about faces such as pictorial information (i.e. lighting or pose) and structural information (to distinguish the face from other faces). In their model, recognition of familiar faces is based upon matching between products of structural encoding and previously stored structural codes in ´face recognition units` describing the face. However, usually witnesses in the real-life investigations are describing unfamiliar faces. Indeed, Bruce and Young suggested that there are meaningful differences in the processes underlying the recognizing of (i) familiar and (ii) unfamiliar faces. Familiar face recognition, they said, is based more on structural encoding, face recognition units, person identity nodes, and name retrieval, whereas unfamiliar face recognition is dominated by pictorial codes such as details of viewpoint, expression, and lighting. The implication of their suggestion is that suitable tools, for instance feature checklists, are needed to aid the recollection of details about unfamiliar faces.

An example of a feature checklist designed to help to gather descriptions of faces is Aberdeen Face Rating Checklist (Shepherd & Ellis, 1996; Meissner, et al., 2006) which consists of 50 items on which witnesses are asked to rate individual features on five point scales. According to these researchers the checklists seem to produce information about faces which the witnesses would not remember without using such an aid. On the other hand, Wogalter (1991, 1996) showed that such feature checklists can produce incorrect features (e.g. by mentioning features which were not originally encoded about the face) and interfere with witnesses' ability to identify the perpetrator. A major mechanism related to interference between description and recognition of faces is the concept of *verbal overshadowing* which is now described more fully.

1.7. VERBAL OVERSHADOWINGS AND VERBALIZATION

It is known that recalling a face is harder than recognizing it (Shepherd & Ellis, 1996). People have considerable experience of recognizing faces, especially familiar faces, but are less experienced at describing faces. Wells (1985) found that distinctive faces tended to be both easier to describe and to recognize than less distinct faces (see also Newel, Chiroro, & Valentine, 1999). Therefore, if a culprit is ´just a usual face from the crowd` it could be harder to remember him/her than if he/she were distinctive.

Although being separate processes, giving verbal descriptions of perpetrators may influence visual identification performance. This effect is referred to as ´verbal overshadowing` (Schooler & Engster-Schooler, 1990) according to which the negative effects of verbalization could be due to a mismatch between the visual information or processes associated with the original experience and the verbal information or processes associated with the act of verbal description.

A meta-analysis by Meissner and Brigham (2001) found a small yet significant verbal overshadowing effect (effect size $r = .12$) demonstrating that overall participants who described a target face were 1.3 times more likely to later misidentify the face from a lineup than those who did not describe a face. In studies where participants were forced to generate rather elaborate descriptions of faces, and were later asked to identify these individuals in a lineup identification task, such elaborate verbal descriptions led participants to generate inaccurate details which then impaired their recognition performance (Meissner, Brigham, & Kelley, 2001). Meissner (2002) found that such an effect was maximized when participants were forced to provide elaborate descriptions of the face.

Wells, Charman, and Olson (2005) contended that causing participants to recall incorrect details, which then distracts the original memory of the face, is the likely cause of verbal overshadowing effects (Schooler & Engster-Schooler, 1990, termed this process ´recoding interference`). Schooler and Engster-Schooler described the recoding interference as the situation in which "the verbalization of a visual memory can foster the formulation of a nonveridical verbally biased representation corresponding to the original stimulus" (p. 62). This partly explains why the overshadowing effect is observed mainly with non-verbal stimuli (such as faces) which are difficult to describe verbally, but not with stimuli that are more easily described.

There are some other investigatory procedures which can influence person recognition. For example, building a face composite seems to diminish the likelihood that a person will later be able to identify that face from a lineup (Wells et al., 2005). This effect could be considered to be similar to the verbal overshadowing effect because witnesses have to transform the visual information into a verbal mode to complete the composite and this is usually even more detailed than most verbal descriptions (and the chances of participants inserting features that they do not recall is high). Also, some studies (Brown

& Lloyd-Jones, 2003; Dodson, Johnson, & Schooler, 1997) have found that describing one face can impair recognition memory for other faces. Therefore, it is important to find a balance between the tasks of face description and recognition to prevent the occurance of interference effects and also to maintain accuracy in those two tasks.

Not all researchers have found a verbal overshadowing effect. Mauldin and Laughery (1981) found that after verbal descriptions their participants' later accuracy in face recognition actually increased. A series of studies by Clifford (2003) found no evidence of verbal overshadowing at either short or longer delays (15 minutes or one week later) for both adults and children (similarly to Memon & Rose, 2002; Memon & Bartlett, 2002). Lyle and Johnson (2004) also found that verbalization did not affect identification accuracy of the target face (i.e. they found no effect of verbal overshadowing). In their work, verbalization of the target face actually increased the accurate rejection of the distractor face which indicates that verbalization has a rather positive effect of reducing false alarms to the distractor. One possible explanation of this finding is that providing descriptions of a target face may increase the difference between the memorial representation of the target and distractor faces, thereby increasing the likelihood that any recalled cue is specific to one of the faces. However, the participants may have difficulties discriminating whether the cue is specific to the target or to the distractor faces (Johnson, Hashtroudi, & Lindsay, 1993).

Thus, Lyle and Johnson (2004) suggest that verbalization may affect face recognition by creating a mismatch between the ways the described faces are encoded versus retrieved. They state that faces are assumed to be encoded and retrieved holistically but (as stated above) verbalization induces people to ´process` faces by retrieving featural information (information favoured in verbal descriptions). Therefore, target faces could be encoded by features when the participants know that they will be required to prove verbal descriptions but holistically when no verbal description is known to be required.

1.8. EMPIRICAL RESEARCH WITH CHILDREN

Even very young children can provide some descriptive details (Sporer, 1996). Also, even the youngest of child witnesses are capable of accurately reporting the behaviour of others (Ceci, Crossman, Gilstrap, & Scullin, 1998) and therefore their testimonies are

becoming more admissible in justice systems (when they are interviewed in a developmentally appropriate manner).

Regarding verbal recall (which may contain descriptive details), Dekle, Beal, Elliott, and Huneycutt (1996) found that compared to adults, children's free recall is less complete (but equally accurate). When children are allowed to recall information freely or through the use of general questions, even very young children can produce material that is as accurate as that given by adults (Hutcheson, Baxter, Telfer, & Warden, 1995). In general, children's testimonies contain significantly fewer details than adults' or adolescents' but the accuracy of the information can be comparable (Marin et al., 1979; Davies & Flin, 1988). The fact that adults' statements are usually longer and more detailed than children's (Marin et al., 1979; Dent & Stephenson, 1979; Davies, Tarrant, & Flin, 1989) could be due to younger children not encoding and storing information as effectively as adults and older children do (Brainerd, Reyna, Howe, & Kingma, 1990). Lamb et al. (2003) found that more details were elicited from older than younger children in response to all types of prompts, but there were no age differences in the proportion of details (about 50%) elicited using invitations.

Preschoolers often report relatively little information in response to open-ended invitations for free recall (Poole & White, 1991). Prompting children's recall can occasion fuller reporting, but some of this will be inaccurate. Younger children (nine years old and less) tend to focus more on the exterior rather than interior facial features of (familiar) others (Davies et al., 1989). Adults usually describe well the clothing of the perpetrator, children describe the actions, and adolescents describe the appearance (face, body, height, weight) (King & Yuille, 1986; Cesniene & Bandzeviciene, 2005). Children's preference for actions could be explained by the notion that when they learn language, they first learn the actions associated with frequently heard verbs (Glenberg & Kaschak, 2002). Therefore, they could more notice (and comment on) what a person *does* than how the person looks. When remembering real-life events, children have been found to be less likely to freely recall descriptive information than central actions and objects (Tobey & Goodman, 1992).

Pozzulo and Warren (2003) found that adults were better describers both quantitatively and qualitatively than were 10 to 14-year-old youths. Adults reported more features of the face and body (height, weigh and build), race, and clothing, whereas youths reported

various accessories. Youths were less accurate than adults in describing interior facial features (e.g. nose), the age, and body characteristics.

1.8.1. Height, weight, and age

Although children's accounts are generally accurate, their estimates about the stable characteristics (such as height, weight, and age) of a person can be quite poor. Younger children's limited performance with many descriptors could be due to their limited experience of (or exposure to) different persons (see Davies, 1996). Davies, Stevenson-Robb, and Flin (1988) found that seven to twelve year old children were poor at estimating the height, weight and age of an unfamiliar person, especially age. Goetze (1980, see in Davies, 1996) found increases with age in the accuracy of estimates of height, weight, and age in 8, 11, and 13-year-old children.

Overall, studies comparing actual and estimated height (Janssen & Horowski, 1980) have revealed that the estimations of older children were more correct than of younger children. Also, children's own height probably makes it difficult to estimate the height of adult persons. In the case of age estimates, the problem appears to lie in lack of knowledge among younger subjects of the appropriate facial cues to aging (Ellis, 1990) or the concept of aging itself.

1.8.2. Gender of witness

Cesniene and Bandzeviciene (2005) found that girls were more accurate in answering open-ended questions about the observed event than boys (see also Kask, Bull, Heinla, & Davies, 2007). Also, girls were more correct than boys in describing people and actions.

To conclude, even young children's accounts can be as accurate to those of adults but not so rich in details. Children tend to focus their attention on different characteristics than adults do. Studies suggest that a developmental trend may be present in that with increasing age, interior features of faces are more noted by children although they may still be more difficult to describe, because they require a richer vocabulary.

1.9. CHILDREN'S FACE RECOGNITION

Despite development of an early preference for facial stimuli (Slater, Bremner, Johnson, Sherwood, Hayes, & Brown, 2000), children often seem to perform relatively poorly on face recognition tasks until adolescence (Taylor, Edmonds, McCarthy, & Allison, 2001). As stated above, the ability to recognise faces increases with age (Ellis, 1990). Chung and Thomson (1995) in their review noted that children's ability to later recognize once-presented photographs of unfamiliar faces improves particularly between the ages of five and ten years.

Carey and Diamond (1977) suggested that before the age of eight, children recognise a face by its individual features (piecemeal approach), and only after that age do they switch to a more holistic recognition strategy. It has been found that children aged four to ten years categorize faces by focusing on single facial features, whereas the majority of adults base their decisions on the whole face (Schwarzer & Massaro, 2001). Thus, one possible explanation for children's poorer face recognition performance could be that their development of face expertise takes at least ten years to acquire (Carey & Diamond, 1977; Diamond & Carey, 1977; Diamond & Carey, 1986; Chung & Thomson, 1995). The 'encoding switch hypothesis' suggests that if young children encode faces in a piecemeal fashion, then disrupting holistic information (which they do not process) should produce less interference for the younger than for older children. This notion could be relevant in processing of different race faces where own-race faces would be processed in a more holistic manner whereas other-race faces would be processed more in a piecemeal approach. Thus, the own-race superiority effect may actually be less for younger children.

However, Tanaka, Kay, Grinnell, Stansfield, and Szechter (1998) claimed that children recognize faces holistically by the age of six years (also see Pellicano & Rhodes, 2003). Thus, their results fail to support the notion that children switch from a featural encoding strategy to a holistic strategy across the primary school years. Therefore, it is not clear yet at which point in children's development this shift in processing is occurring.

Developmental differences have also been found regarding children's whole person recognition. For example, Seitz (2002) found that person recognition (i.e. the percentage

of correct identifications) became more accurate between eight years and adulthood. Her results showed that visual information processing of both face and whole person recognition was similarly holistic for children and adults (i.e. no developmental shift appeared). She concluded that whole person recognition does not rely on different processes from face recognition for eight- and ten-year-olds as well as for adults. However, it has been noted that the inner face advantage for adults' recognition of (familiar) faces is not in evidence for children until the age of 15 (Campbell, Coleman, Walker, et al., 1999). Furthermore, the literature on neuroimaging points toward selective activation of different cortical regions when (a) images of the human body and (b) images on the human faces are processed (Downing, Jiang, Shuman, & Kanwisher, 2001).

Bruce and Young (1986) put forward three hypotheses for understanding the effects of (adult) participant age on the false recognition of unfamiliar faces. These could be valid also for children. Firstly, a ´compensation hypothesis` that relates to greater limitations in older adults' (60 years and older) encoding and retrieval processes which might lead them to have a looser criterion for recognizing faces and show reduced discrimination between familiar and unfamiliar faces. (As younger children's cognitive processes are still developing, such compensation might also influence their face recognition.) Secondly, a ´resemblance hypothesis` according to which older adults experience more resemblance in response to new (than to familiar faces) than do younger adults (because of their greater past experience regarding faces.) (This hypothesis could be valid also for children but in an opposite direction - they lack extensive experience of face recognition.) Thirdly, a ´context recollection hypothesis`. Older adults might fail to recall contextual information about the face because of their diminishing memory capacities and therefore rely on a feeling of familiarity for the face. (This could be valid also for (younger) children due to their, as yet, poorer memory capacities.)

1.9.1. Own-age bias

Own-age biases in adult face recognition have been found. Bäckman (1991) found that young and older adults recognized better faces of the own age group. An own-age bias has been found recently regarding children's face recognition (Anastasi & Rhodes, 2005), which suggests that children may encode faces close to their age differently from adult faces. Children's encoding methods were mentioned above but even more

important is how children express what they have seen or perceived. Therefore, now issues concerning questioning methods (e.g., retrieval) will be covered which is a crucial point in police investigation.

1.10. QUESTIONING METHODS WITH CHILDREN

As noted above, children freely recall less information than adults. Therefore, it is very important how interviewers question them. For the optimum result it is important for the interviewers to know appropriate interviewing methods and also their impact on the quality and quantity of information (the issues on interviewing methods are more in depth covered in Chapter 3).

Asking children too many specific questions can lead to an increase in errors (Ceci & Bruck, 1993; Greenstock & Pipe, 1996; Hutcheson et al., 1995; Marin et al., 1979) which could result in providing an erroneous person description. Poole and White (1991) found that while six and eight-year-old children were as accurate as adults when responding to open-ended questions, repeated specific questions led to decreased accuracy for children of all ages, whereas repeated general questions had no such deleterious effect (see also Marin et al., 1979). This notion is supported also by Memon and Vartoukian (1996), who found that seven-years-old children were accurate on both open and closed questions. Dent and Stephenson (1979) found that 10 to 11-year-old children produced the largest amount of information in response to specific questions, but this was also accompanied by the highest rate of error. Therefore, with some questioning methods the amount of information increases in children's accounts, whereas the accuracy decreases which results in having longer but more erroneous person descriptions. Also, age differences in children and their verbal abilities have a strong effect on children's recollections (see Chapter 3). To aid children to recall more information about the person/event structured interviewing methods have been successful such as the NICHD protocol (Orbach, et al., 2000), Achieving Best Evidence (Home Office, 2002), or cognitive interview (Fisher & Geiselman, 1992, see Chapter 3).

Cassel, Roebers, and Bjorklund (1996) found that children under the age of seven are more vulnerable to strongly misleading questions. During cross-examination style questioning both five to six-years-old children (Zajac & Hayne, 2003) and 9 to 10-year-old children (Zajac & Hayne, 2006) have been found to make frequent changes to their

original responses. Younger children were just as likely to change a correct response as they were to change an incorrect one, but older children were more likely to change incorrect responses than correct ones (but nevertheless changing over 40% of their correct responses).

Some other factors might influence children's erroneous person descriptions. For example, with *delay* (between seeing a person and giving verbal description about him/her) studies have shown high inaccuracy rates for even open-ended questions in children's recollections (Poole & White, 1993; Pipe, Sutherland, Webster, Jones, & La Rooy, 2004; Jones & Pipe, 2002, see also La Rooy, Pipe, & Murray, 2007). Repeated interviewing heightened *misinformation effects* only when children received the two interview session temporally close to the event and memory test (Melnyk & Bruck, 2004). Odinot and Wolters (2006) found that longer *retention intervals* before recollection resulted in lower accuracy, whereas repeated recall had little effect on accuracy. Similarly to adults, *a weapon focus effect* (concentration of witness's attention to a weapon which impairs the ability to remember other details about the person) have also been found recently in children (Davies, Smith, & Blincoe, 2008; Pickel, Narter, Jameson, & Lenhardt, 2008).

1.11. CONCLUSION

Improving person descriptions has crucial value. Suppose a police officer, soon after a crime has been committed, has to make a quick decision about the culprit based on the person description information given by a child witness. If the estimates of age, height, weight differ a lot from the actual characteristics of the perpetrator, a lot of police time would probably be wasted and the culprit could disappear without difficulty.

Translating a face into a verbal description is quite a difficult task (Sporer, 1996), and also recalling a face is harder then recognizing it (Shepherd & Ellis, 1996). There is probably no strong relationship between face description quality and identification accuracy, especially for children (e.g. Sporer, 1996; Wells, 1985). Children's identification accuracy is examined in Study 1 (see Chapter 2) in this thesis.

However, giving verbal descriptions of perpetrators can interfere with identification performance and overshadow visual memory – the ´verbal overshadowing` (Schooler &

Engster-Schooler, 1990). In descriptions adult witnesses mention more general features such as race, age, height, stature, and also clothes and face. Most of these features are important in person recognition and they can not be easily manipulated or changed (except clothes).

The information about persons to be gained from very young children is generally likely to be less complete but as accurate as their older counterparts and adults (Marin et al., 1979; Davies & Flin, 1988). While adults usually describe well the clothing of the perpetrator, children focus more on actions (King & Yuille, 1986). Children's estimates about the stable characteristics (such as height, weight, and age) of a person can be quite poor.

Although the ability of recognise faces increases with age (Ellis, 1990), children often perform relatively poorly on face recognition and encoding tasks until adolescence (Taylor, et al., 2001). Children's and young adults' (different race) face recognition performance is examined in Studies 4 and 5 (see Chapters 5 and 6).

As the children's free recall can be limited, it is important to ask them developmentally appropriate questions, such as open-ended questions (compared to closed questions) which may produce less information but increase accuracy (Hutcheson et al., 1995). The interviewers questioning methods with children are examined more closely in Study 2 (see Chapter 3).

As younger children's problems with many descriptors could be due to their limited experience (see Davies, 1996), it has been suggested that providing children with possible ranges or specific anchors (Sporer, 1996; Meissner, Sporer, & Schooler, 2006), or a colour plate, may lead to better results for some aspects of person descriptions. This topic will be returned later in Study 3 (see Chapter 4).

2. CHAPTER 2: A META-ANALYSIS OF IDENTIFICATION STUDIES INVOLVING CHILDREN.

2.1. ABSTRACT

Purpose. The identification accuracy of children and adults was examined in a meta-analysis. A number of relevant studies have been published since the last meta-analysis by Pozzulo and Lindsay in 1998, therefore an up-to-date review and meta-analysis is now required.

Methods. The sample consisted of 28 different studies with a total of 4477 participants (children=3169; adults=1308). These data were derived from 26 published studies, one conference paper, and one unpublished manuscript.

Results. There were no significant effects of age (i.e. children versus adults) on correct identifications and false alarms, although children overall made twice as many false alarms than adults. Children's identifications were more accurate when there were fewer members in the lineup. Delay had no effect on children's correct identifications and false alarms. For showups, children's correct identifications were higher and false alarm rate lower than for simultaneous or sequential lineups. There was no effect of the mode of lineup presentation (live, photo or videospreads) for children.

Conclusions. The results demonstrate that the differences between children and adults in correct identifications seem to be diminishing. However, new presentation methods for children are needed to decrease their inability to correctly reject target-absent lineups. Some future directions for identification research are presented.

2.2. INTRODUCTION

Historically children have been viewed by the legal system as less reliable and less accurate than adults (Ceci & Bruck, 1993). One important aspect of witnesses is their ability to identify the perpetrators of crime. Compared to the number of studies concerning children's recall of what happened during an event, there are fewer studies of children's person identification performance. Even fewer studies have directly compared children's and adults' identification ability. In 1996 Davies reviewed studies of children's identification abilities and in 1998 Pozzulo and Lindsay published a meta-analysis. A number of relevant studies have been published since 1998 and therefore an up-to-date review and meta-analysis is now required.

In the last decades an increasing number of studies have been published concerning eyewitness' *identification* abilities. *Identification studies* are usually constructed within the context of applied settings about factors that may influence eyewitness testimony. In eyewitness studies, participants are normally exposed to an 'event', either live or in videotape format, usually without prior knowledge that they have to identify the person(s) involved sometime later. Typically, there is only one target and the recognition task takes place very soon after the event. However, in some studies there have been longer delays between first seeing and later trying to identify the target.

In contrast, *face processing* studies are designed to answer theoretical questions about the processes by which faces are recognized. In such studies participants typically see a large series of photos of target faces (e.g., 20) and are usually told at that time to remember the faces for a subsequent test. Participants may be tested immediately after seeing the faces or after a delay of days, weeks, or even months. For the recognition phase, photographs of the many previously seen faces are mixed in with a large sample of new faces and the resultant large set is shown to participants whose task is to identify all the previously seen photos.

In their meta-analysis of *facial identification studies*, Shapiro and Penrod (1986) reported that adults yielded a large effect size for hits than did children. However, King and Yuille (1986) found that when the suspect was present in the lineup, no age differences were present (more than 60% of subjects of all ages made correct choices). Chance and Goldstein (1984) in their face processing study reported hit rates between 35% and 40% for 4 to 5 year-olds; 50% to 58% for 6 to 8 year-olds; 60% to 70% for 9

to 11 year-olds; and 70% to 80% for 12 to 14-year-olds. Adult performance has been shown to be similar to that of 12 to 14-year-olds (Goldstein, 1977).

Pozzulo and Lindsay (1998) in their meta-analysis found that the proportion of hits and level of accuracy increased with participants' age. They found that children of four years of age had significantly lower rates of correct identifications than adults (47% vs 67%); but five to six-year-olds made significantly more correct identifications than adults (71% vs. 54%); nine to ten-year-olds were similar to adults (47% vs 48%), and the 12 to 13-year-olds also had adultlike performance (66% vs 57%). These authors contended that the later onset of adult-like hit rates in face processing studies than in eyewitness studies was due to the greater number of targets in the former type of study.

King and Yuille (1986) found that target-absent lineups produced a strong age effect where only 15% of the eight to nine-year-olds compared to 39% of the 10- to 11-year-olds and 58% of the 12- to 14-year-olds correctly rejected the lineup. A similar pattern was present in Davies et al. (1988).

It is important for witnesses to identify the perpetrator but it is also important for witnesses to correctly reject a lineup when the target is not present (e.g. when an innocent suspect has been accused of the crime). In *face processing studies*, false positives have been found to decrease with age (Chance, Turner, & Goldstein, 1982). Indeed, Shapiro and Penrod (1986) found that one of the largest effect sizes of age was for false positives. Chance and Goldstein (1979) reported that 13-year-olds (and older) produced a similar false-positive rate as adults.

2.2.1. Factors moderating identification accuracy

There are a variety of factors that could influence *identification* performance. The method sections of relevant studies were examined to note any variables possibly influencing accuracy such as method of lineup presentation, type of lineups, mode of target presentation, delay, and the number of persons in the lineup.

2.2.2. Target-present and target-absent lineups

Studies investigating the children's identification capabilities have shown that children above six years of age usually perform as well as adults in terms of the number of correct identifications from target-present (TP) lineups, but when tested with target-absent (TA) lineups, children's performance is poorer (Parker & Ryan, 1993; Gross & Hayne, 1996). In particular, children make more false choices (Parker & Ryan, 1993). According to Shapiro and Penrod (1986), even a 'not here' option may not reduce this tendency.

Pozzulo and Lindsay (1998) found that preschoolers were less likely to correctly reject a TA lineup than adults (39% vs 98%). Surprisingly, young children (5 to 6-year-olds) were similar to adults with their correct rejection rate (57% vs 65%). Both older children (9 to 10-year-olds) and adolescents were significantly less likely to correctly reject TA, lineups than adults (41% vs 70%) and (48% vs 74%). Older children (9 to 10 years) were also less likely to make correct rejections than adults. The current review examines whether the type of lineup (i.e. TP and TA) has an effect on identification accuracy for children.

2.2.3. Method of lineup presentation

A lineup generally takes one of three forms: simultaneous lineup, sequential lineup, and showup. In a **simultaneous lineup** all the lineup members are presented to the witness at the same time. One of the criticisms of this method is that witnesses can use a 'relative judgment' strategy (i.e. where they choose the person who most looks like the target, even though the target may not be present, Wells, 1984). In simultaneous target present lineups children have been found to show a significant tendency to guess (Lindsay, Pozzulo, Craig, Lee & Corber, 1997; Beal, Schmitt & Dekle, 1995). Dekle et al. (1996) not only found that children were more likely than adults to identify the perpetrator correctly when the target was present, but also for TA lineups to make more incorrect identifications. For simultaneous lineups Pozzulo and Lindsay's (1998) meta-analysis found that nine to ten-year-olds were less likely than adults to correctly reject a TA lineup (46% vs 62%).

Sequential lineups are thought to diminish the effect of using a relative judgment strategy (see Lindsay & Wells, 1985). In a sequential lineup, the faces are shown to the

witness one at a time and the witness has to make a decision for each whether or not it is the target. Witnesses are informed that they cannot again see any previously seen faces in the set, and also they cannot change their decision once it has been made. Furthermore, if they choose a face they are not shown the remaining faces in the set. Also, witnesses are blind to the number of faces in the set lineup. In their pioneering study Lindsay and Wells (1985) found adults made significantly fewer false alarms in the sequential lineup condition (18.3%) than in the simultaneous lineup condition (35.0%) when the target was absent. [When the target was present in the sequential lineup condition, correct identifications (50%) were not significantly lower than in the simultaneous condition (58%)].

Steblay, Dysart, Fulero, and Lindsay's (2001) meta-analysis of 23 studies with adults found that for TP lineups there were more correct identifications and fewer incorrect rejections from simultaneous lineups and no significant difference between false identifications from sequential or simultaneous lineups. On TA lineups, sequential presentation increased the number of correct rejections and decreased the number of false identifications. Other studies have also found that sequential lineups tend to have no effect on adults' correct identifications but reduce false identification rates in target absent arrays (Lindsay, Lea, Nosworthy et al., 1991; Searcy, Bartlett, & Memon, 2000). According to Wells, when showing witnesses a sequential lineup, witnesses make an 'absolute judgment' for each face by comparing their memory about the target with that face. Also, as they do not know how many faces there are in the lineup, they adopt a more stringent criterion (than for simultaneous lineups) for an identity match.

However, although sequential lineups increase adult correct rejections, they do not seem to do so with children (Lindsay et al., 1997). Pozzulo and Lindsay (1998) in their meta-analysis found that nine to ten-year-olds were less likely to correctly reject TA sequential lineups than adults (.21 vs .81). Although children were less likely than adults to correctly reject TA simultaneous lineups, the difference was even greater when sequential lineups were used.

A third alternative task is a **showup** – the presentation of a single person to the witness for possible identification. Witnesses may use an absolute judgment in a showup but this procedure may be highly prejudicial to the suspect (Lindsay, et al., 1997; Yarmey, Yarmey, & Yarmey, 1994). In a showup witnesses know who the suspect is and this

knowledge may influence their decisions, especially increasing the possibility of false identifications (Lindsay et al., 1997; Yarmey & Yarmey, 1997). A meta-analysis of adult performance by Steblay, Dysart, Fulero, and Lindsay (2003) which included showup presentation mode revealed similarly that in TP conditions, showups and lineups had equal in hit rates, but in the TA condition, showups produced a higher level of correct rejections. However, false identifications were more numerous for showups in which the innocent suspect more strongly resembled the perpetrator.

To sum up, five-years-old and older children do not seem to significantly differ from adults with respect to correct identifications (Pozzulo & Lindsay, 1998). However, there may be a greater tendency for them to make a choice from TA lineups than correctly reject the lineup. The present review examines whether the method of lineup presentation impacts on the identification accuracy of children compared to adults.

2.2.4. The mode of target presentation

Participants in eyewitness research may be exposed to different modes of target presentation that is live, photo, or a video/DVD presentation. Few studies have examined the effect of different modes of lineup presentation on identification accuracy. Lindsay and Harvie (1998) examined identification rates of a videotape and a slide sequence of a staged crime (presented to individual participants), and the same event presented live (either to individuals or to large groups). They found that correct identification rates were not affected by mode of target presentation. However, correct rejection rates were higher in the slide and video conditions, compared to the live exposure conditions. One explanation to this finding could be that in real life the range of perpetrator characteristics to be perceived is larger (and therefore harder to encode effectively) compared to slide or video presentation.

Shapiro and Penrod (1986) in their meta-analysis found an effect of mode of target presentation, in that when the event was more realistic (live or on videotape) people remembered the target more accurately than when viewing it via photographs, slide show or drawings. Lindsay and Pozzulo (1998) in their meta-analysis found that with a live target presentation, children rejected TA lineups significantly less often than did adults (53% vs 77%), and that 9 to 10-years-old children were also less likely to make correct rejections than adults compared to when the target was presented as slides (25%

vs 62%). There was a smaller difference between adults and children when live exposure was used. The present review examines whether the mode of target exposure has an effect on children's identification abilities.

2.2.5. Delay

Post-event factors such as time delay (between seeing a person and later identifying him/her) may also have an impact on both false and correct identifications. Davies (1996) reported that delay has a stronger effect on children's errors on TA lineups than on TP lineups. However, when Peters (1987) tested children's memory of their autobiographical experiences (persons whom they have seen etc.) he found that when recognizing persons they made many more errors on TA than TP lineups, but there was no effect of delay. Goodman, Bottoms, and Schwartz-Kenney (1991) and Peters (1991) also found no effect for delay. However, perhaps a three to four week delay between seeing the target and trying to recognize him/her is not sufficiently long (see also for adults Shepherd, 1983).

Behrman and Davey (2001) did find an effect of delay for showups (accuracy rate dropped substantially after seven days). Cain, Baker-Ward, and Eaton (2005) found that older preschoolers were not affected by delay, whereas toddlers had greater difficulties after a delay of three months. In this review the effect of delay on children's (and adults') identification abilities is examined.

2.2.6. Number of members in the lineup

Previous meta-analyses seem not to have examined the possible effect of the number of members in the lineup on identification performance. This review examines this factor.

2.2.7. Aims of the meta-analysis.

The present meta-analysis of eyewitness *identification studies* examines the rates of correct identification by different age children and adults, and is especially interested in at what age children's correct identification reaches the same level as adults'. The effects of method of lineup presentation and mode of target presentation on children's identification accuracy are examined. Concerning the method of lineup presentation, it is hypothesized that the differences are smaller between adults and children regarding

TP than TA lineups (especially in comparing simultaneous and sequential lineups). Also, novel factors in meta-analysis concerning child witnesses are examined (i.e. whether time delay and the number of persons in the lineup have effects on children's identification accuracy). It is hypothesized that time delay has a negative effect on the identification accuracy and having less number of persons in the lineup increases children's identification performance. Finally, random effects regression analysis was used in this meta-analysis to provide more comprehensive results compared to fixed effects analysis.

2.3. METHOD

2.3.1. Sample

Computer search using relevant databases such as the EBSCO PsycInfo and PsycArticles was conduced for the period 1979 to February 2006 using keywords such as *eyewitness; witness; children; identification; lineup; face recognition*. Publication bias is a common criticism of meta-analysis, namely the results of a meta-analysis may be artificially inflated due to the tendency to publish significant findings only which is commonly known as the file drawer problem (Rosenthal, 1991). Therefore, every effort was made to obtain all relevant studies. Via keyword search nearly 250 papers were obtained. The studies were analyzed and those which were not relevant to forensic/legal settings were excluded. A few studies have also presented participants with several lineups generating possible practice effects (e.g. Leippe, Romanczyk, & Manion, 1991; Parker, Haverfield, & Baker-Thomas, 1986). These studies were also not included in the present review.

Papers retained for this meta-analysis had to meet several inclusion criteria. One criterion was straightforward data, i.e. the paper needed to include an effect size (e.g., *r, t, F,* chi-square) or report enough data to compute one (e.g., means and variances across two groups). Secondly, studies included in this meta-analysis were restricted to those which were conducted in the field of forensic/legal psychology and involved children (and an adult comparison group). The age of the child sample was not restricted and the age range for child participants was 3 to 15 years (adults' age was 18 years and older). Direct contact with lineup researchers who have published in the area (n = 23) provided access to additional relevant published and unpublished work. In any study with more than one child sample and only one adult sample, the adult sample was used for

comparison with each child sample. (Parker et al. (1986) used both adult and child targets but because studies involving child targets are very rare, this review includes only studies with adult targets.) The studies included in this review are indicated with an asterisk in the reference list.

The final sample consisted of 28 different studies with a total of 4477 participants (children=3169; adults=1308). These data were derived from 26 published studies, one conference paper, and one unpublished manuscript. For target-present lineups there were 1882 children (2 to 344 per study) and 744 adults (12 to 125) drawn from 27 experiments (involving 70 comparisons), and for target-absent lineups, 1287 children (2 to 309) and 564 adults (12 to 119) drawn from 18 experiments (involving 49 comparisons).

2.3.2. Study characteristics

Information concerning six characteristics was coded for each study.

1. *Age.* Five age groups were identified across the studies (kindergarten M= 3 to 4 years; preschoolers M= 5 to 6 years; young children M= 7 to 8 years; older children M= 9 to 10 years; and adolescents, M= 11 to 15 years).

2. *Presence versus absence of target.* Lineups were classified as either target-present tasks (the target was in the lineup) or target-absent tasks (the target was not in the lineup shown to the participants). Twenty seven studies used target-present lineups and eighteen used target-absent lineups.

3. *Method of lineup presentation.* If all lineup members were presented at the same time, this was coded as a simultaneous lineup. If a witness was shown faces one at a time and asked to make a decision as to whether or not it was the target after each presentation, it was coded as sequential lineup. Where only one photo or person was shown this, it was coded as a showup. Of the target-present lineups, 85% were simultaneous, 8% sequential, and 7% were showups; of the target-absent lineups were 74% simultaneous, 16% sequential, and 10% showups.

4. *Delay.* The delay between exposure and identification was noted; for target-present lineups the range was from zero to four weeks (for 73% of participants the delay was less than 30 minutes) and for the target-absent lineups ranging from zero to 36 hours (for 94% of participants the delay was less than 30 minutes).

5. *Number of members in the lineup.* The number of members in the lineup was noted (usually six persons). In 90% of the target-present lineups, there were six persons, 4% eight persons, 3% five persons, and 3% four persons. In 92% of the target-absent lineups there were six persons, 4% eight persons, and 4% four persons.

6. *The mode of target presentation.* Three target presentation modes had been used: live, photo or video. For target-present lineups, 32% of the lineups were presented live, 64% via photographs, and 4% via video; for target-absent lineups 26% were live presentation, 72% photographs, and 2% via video.

2.3.3. Dependent variable

The dependent variables used in the meta-analysis were the performance in each of the studies for target present (i.e. correct choosing/'hits') and for target absent (i.e. incorrect choosing/'false alarms'). (Since for target present incorrect choosing is the converse of correct choosing, this was not included in the meta-analysis).

2.3.4. Statistics

Frequency of correct identifications or false alarms per condition allowed chi-square to be calculated. Therefore, chi-square and random effects regression analysis were used in the present meta-analysis (compared to a fixed effect model in Pozzulo and Lindsay's (1998) meta-analysis).

Some authors suggest it as a preferred strategy instead of fixed effect model because of the generality of the random effects model (Mosteller & Colditz, 1996). A random effects regression analysis was used because this form of analysis assumes that each observed effect size differs from the population mean by subject-level sampling error

plus a value that represents other sources of variability assumed to be randomly distributed (Lipsey & Wilson, 2001). A fixed effects model can under certain circumstances create problems such as decreasing the number of degrees of freedom which may increase standard error (Mundlak & Yahav, 1981). These authors also point out that using a random effects model can eliminate cross-sectional variance in the independent variables which again can increase standard error (and may make some standard errors infinite), and finally, exacerbate problems in measurement error.

As stated above, one of the concerns with regards to meta-analytic reviews is the file drawer problem (Rosenthal, 1991). More specifically, studies that reject the null hypothesis (i.e., produce statistically significant results) are more likely to be published. Therefore, if meta-analysts draw their sample only of studies from the published literature, then there is an inherent bias where the stronger and more significant effects are likely to be included which could lead to a biased conclusion of the magnitude of the relationship involved. Rosenthal (1991) developed a solution which addresses this problem at the analytical level and is termed the 'Fail-Safe N' (FSN) statistic. This FSN corresponds to "the number of additional studies required in a meta-analysis that would be necessary to reverse the overall probability obtained from our combined test to a value higher than our critical value for statistical significance" (Wolf, 1986, p. 38).

This statistic provides further validation of the external generalisability of the findings (i.e. if only a few additional studies are required to change the direction of the conclusions, the findings should be viewed as cautionary whereas a FSN of several hundred makes the findings considerably more robust). In this meta-analysis a FSN was calculated to estimate the number of additional tests averaging null results that would be needed in order to bring the significance level attained through the meta-analysis to a $p = .05$. Rosenthal (1984) suggested using a "5k + 10" benchmark for determing a tolerable fail-safe number for a database, where k is the number of hypothesis tests.

2.4. RESULTS

2.4.1. Target present lineups

Firstly analyses for TP and then secondly for TA lineups will be presented. The chi-squared goodness-of-fit statistic, based on model residual variation, indicates that the random-effects multiple regression models provides a good fit to the data for children

$\chi^2(65) = 68.51, p < .36$, and to the combined data for children and adults $\chi^2(23) = 22.79$, $p < .47$.

Age effects

Examination of the seventy four comparisons (see Table 2.1) indicated overall that for correct identifications, the effect of children's age was statistically significant $t(64) = 2.73, p < .01, n = 74$, Cohen's $d = .68$, $r = .32$ with the correct identification rates being for kindergarten children (.46, n=10); preschoolers (.61, n=24); younger children (.59, n=12); older children (.54, n=16); and adolescents (.63, n=12).

Approximately 400 studies (FSN) averaging null results would be necessary to achieve an overall combined probability level of $p = .05$. With this database (n=74) and on Rosenthal (1984) our benchmark would be 380 and therefore the current FSN (which is above this benchmark) is acceptably tolerant for future null results.

There was no effect of children (all ages combined) versus adults regarding correct identifications (children .57 vs adults .56, $n = 32$, Cohen's $d = .43$, $r = .21$). The chi-squared goodness-of-fit statistic, based on model residual variation, indicates that the random-effects multiple regression models provides a good fit to the combined data for children and adults $\chi^2(25) = 25.54, p < .43$.

The number of members in a lineup.

For children the number of members in the lineup had a significant effect $t(64) = -5.40$, $p < .01$, $n = 74$, Cohen's $d = 1.35$, $r = .56$, indicating that when there were fewer members children's correct identifications were higher (for 4-member lineup the correct identification rate was .91, for 5-member lineup .66, for 6-member lineup .57, and for 8-member lineup .33). The FSN (755) was acceptably tolerant for future null results. For correct identifications there were no age differences between children (all ages combined) and adults in the number of members in a lineup.

Table 2.1. *Target-Present Hypothesis Tests.*

Year	Author	NoChild	ChildAge	Child Hits	NoAdult	Adult Hits	Lineup Type	Exposure Type	Persons	Delay
1979	Marin	24	5-6	.54	24	.54	sim	live	6	0
		24	7-8	.46			sim	live	6	0
		24	11-15	.75			Sim	live	6	0
		28	5-6	.54	25	.43	sim	live	6	28
1986	Goodman	16	5-6	.94	16	.75	sim	live	5	4
		16	3-4	.38			sim	live	5	4
1986	Parker	24	7-8	.58	24	.71	sim	photo	6	0
1989	Parker	12	9-10	.33	12	.08	sim	photo	6	0
1991	Goodman, Bottoms	23	3-4	.80			sim	photo	6	21
		47	5-6	.52			sim	photo	6	21
1991	Goodman, Hirschman	20	3-4	.30	19	.58	sim	live	6	28
		28	5-6	.54	25	.43	sim	live	6	28
		9	7-8	.77			sim	photo	6	3.5
		16	7-8	.56			sim	photo	6	6
		16	7-8	.25			sim	photo	6	6

1991	Leippe	14	5-6	.79	15	.93	sim	live	6	0
		16	9-10	.63			sim	live	6	0
1993	Parker	12	9-10	.42	12	.42	sim	photo	6	0
		12	9-10	.25	12	.08	seq	photo	6	0
1995	Beal	16	7-8	.56			sim	photo	6	0
		20	7-8	.45			sim	photo	6	0
		16	5-6	.62			sim	photo	6	0
		20	7-8	.80			sho	photo	6	0
1996	Gross	17	5-6	.95			sim	photo	4	1.5
		17	5-6	.87			sim	photo	4	1.5
1996	Ricci	16	5-6	.62			sim	photo	6	0
		24	5-6	.42			sim	photo	6	0
		8	5-6	.75			sim	photo	6	0
		8	5-6	.25			sim	photo	6	0
1996	Schwartz	24	3-4	.37			sim	photo	6	0
		24	5-6	.50			sim	photo	6	0
		24	5-6	.31			sim	photo	6	14
		24	9-10	.43			sim	photo	6	14
1997	Lindsay Exp.1	21	9-10	.71	31	.55	sim	live	6	0

		20	11-15	.80			sim	live	6	0
		26	9-10	.65	58	.62	seq	live	6	0
		31	11-15	.71			seq	live	6	0
		22	9-10	.68	30	.50	sho	live	6	0
		18	11-15	.72			sho	live	6	0
	Lindsay Exp.2	19	3-4	.53	20	.80	sim	live	6	0
		19	3-4	.26	40	.45	seq	live	6	0
		19	3-4	.90	20	.85	sho	live	6	0
1997	Pozzulo	37	9-10	.24	25	.52	sim	live	6	0
		54	11-15	.35			sim	live	6	0
1998	Ricci	15	5-6	.57			sim	photo	8	0
1999	Pozzulo	46	11-15	.65	30	.80	sim	photo	6	0
2001	Goodman	17	5-6	.74			sim	photo	6	14
		17	7-8	.77			sim	photo	6	14
2001	Parker	2	9-10	.33			seq	photo	6	0
		3	9-10	.50			seq	photo	6	0
2002	Memon	17	9-10	.82			sim	photo	6	1
2003	Pezdek	31	5-6	.65	31	.81	sim	video	6	1
		31	7-8	.78			sim	video	6	1

2003	Pozzulo	46	11-15	.65	30	.80	sim	video	6	0
		57	11-15	.65	28	.68	sim	live	6	0
2004	Freire	21	3-4	.30			sim	photo	6	0
		35	5-6	.46			sim	photo	6	0
		33	5-6	.54			sim	photo	6	0
		35	7-8	.47			sim	photo	6	0
		22	7-8	.60			sim	photo	6	0
		10	9-10	.61			sim	photo	6	0
		21	9-10	.73			sim	photo	6	0
		22	11-15	.70			sim	photo	6	0
		14	11-15	.73			sim	photo	6	0
		15	11-15	.77			sim	photo	6	0
2004	Weber	344	11-15	.19			sim	photo	8	0
2005	Brewer	55	9-10	.23	125	.31	sim	photo	8	.48
	Fritzley	20	5-6	.60			sim	photo	6	0
		20	9-10	.89			sim	live	6	0

Note. NoCh – total number of children in test. ChildAge – age of child participants. ChildHits – proportion of correct identifications for children. NoAdult – total number of adults in test. AdultHit – proportion of correct identification for adults. Lineup type: sim – simultaneous; seq – sequential; sho – showup. Persons – number of persons in the lineup. Delay – time delay (in days).

The effect of delay

Time delay did not have effects on children's correct identifications. Also there was no differential effect of delay comparing children and adults on correct identifications.

The type of lineup

The type of lineup influenced children's correct identifications $\chi^2(2) = 12.50$, $p < .01$, $n = 32$, Cohen's $d = 1.61$, $\Phi = .62$, in that their correct identification rates for showups were higher than for simultaneous or sequential lineups. The FSN (365) was acceptably tolerant for future null results. Also, children achieved higher correct identification rates than adults for showups (children .79 vs adults .54, n=4) $\chi^2(2) = 28.26$, $p < .02$, $n = 32$, Cohen's $d = 5.40$, $\Phi = .94$, whereas the differences were (and non-significant) smaller for simultaneous (children .57 vs adults .59, n = 24) or sequential lineups (children .47 vs adults .38, n = 4). The FSN (570) was acceptably tolerant for future null results.

The mode of target presentation

There was no effect of mode of target presentation regarding correct identifications for children (the rates for live presentation were (.61, n = 22); for photospread (.55, n = 49); and for video presentation (.69, n = 3)). There was also no children versus adult differential effect.

2.4.2. Target absent lineups

The chi-squared goodness-of-fit statistic, based on model residual variation, indicates that the random-effects multiple regression models provide a good fit to the data for children $\chi^2(40) = 31.65$, $p < .82$, and to the combined data for children and adults $\chi^2(15) = 13.95$, $p < .53$.

For adults, analysis was conducted only for lineup type and the mode of target presentation due to lack of data regarding the other categories.

Age effects

Examination of forty nine comparisons (see Table 2.2) indicated overall that children's age did not have significant effect on false alarms $t(39) = -1.95$, $p = .059$,

n = 49, Cohen's d = .62, r = .30, even though kindergarten children and preschoolers made somewhat more false alarms than younger children and adolescents (false alarm rates were for kindergarten children (.60, n=3); preschoolers (.60, n=15); younger children (.46, n=4); older children (.54, n=14); and adolescents (.47, n=13)). The FSN (245) was not sufficiently tolerant for future null results, therefore further research is needed before any conclusions can be drawn regarding children's age effect on false alarms. Surprisingly, there was no significant differential effect of children versus adults regarding false alarms (.53 vs. 26, n = 22, respectively).

The number of members in a lineup.

The number of members in the lineup did not have a significant effect on children's false alarm rates $t(39) = 1.87$, $p = .07$, $n = 49$, Cohen's $d = .61$, $r = .29$. The FSN (235) was not sufficiently tolerant for future null results. The false alarm rate for 4-person lineups was .68, for 6-person lineups .52, and for 8-person lineups .75.

The effect of delay

There was no effect of time delay on children's false alarm rates $t(39) = 1.91$, $p = .064$, $n = 49$, Cohen's $d = .61$, $r = .29$. The FSN (235) was not sufficiently tolerant for future null results.

The type of lineup

The type of lineup had an effect on children's false alarms $\chi^2(2) = 6.84$, $p < .04$, n=49, Cohen's d=.80, Φ =.37, in that the error rate for sequential lineups (.69, n=38) was higher than for simultaneous lineups (.53, n=36) or showups (.39, n=5). The FSN (314) was acceptably tolerant for future null results. Children made significantly more false alarms than did adults $\chi^2(2) = 12.47$, $p < .01$, $n = 22$, Cohen's $d = 2.25$, $\Phi = .75$, in simultaneous lineups (children .53 vs adults .31, n=14), sequential lineups (.65 vs .18, n=4), and in showups (.40 vs .15, n=4). The FSN (308) was acceptably tolerant for future null results.

Table 2.2. *Target-Absent Hypothesis Tests.*

Year	Author	NoCh	ChildAge	ChildFA	NoAdult	AdultFA	Lineup Type	Exposure Type	Persons	Delay
1989	Parker	12	9-10	.58	12	.33	sim	photo	6	0
1991	Leippe	12	5-6	.33	12	.08	sim	live	6	0
		15	9-10	.13			sim	live	6	0
1993	Parker	12	9-10	.83	12	.58	sim	photo	6	0
		12	9-10	.67	12	.25	seq	photo	6	0
1995	Beal	16	5-6	.94			sim	photo	6	0
		20	5-6	.50			sim	photo	6	0
		16	5-6	.62			sim	photo	6	0
		20	5-6	.40			sho	photo	6	0
1996	Dekle	18	5-6	.61	66	.59	sim	photo	6	0
		18	5-6	.34	50	.38	sho	photo	6	0
1996	Gross	17	5-6	.53			sim	photo	4	1.5
		17	5-6	.82			sim	photo	4	1.5
1996	Ricci	16	5-6	.63			sim	photo	6	0
		8	5-6	.75			sim	photo	6	0
1997	Lindsay Exp.1	25	9-10	.72	29	.34	sim	live	6	0

		21	11-15	.67			sim	live	6	0
		14	9-10	.79	36	.25	seq	live	6	0
		15	11-15	.80			seq	live	6	0
		25	9-10	.40	119	.07	sho	live	6	0
		20	11-15	.40			sho	live	6	0
	Lindsay	9	3-4	.44	20	.00	sho	live	6	0
	Exp.2									
		9	3-4	.78	20	.05	seq	live	6	0
1997	Pozzulo	37	9-10	.62	28	.57	sim	live	6	0
		54	11-15	.55			sim	live	6	0
1998	Ricci	16	5-6	.82			sim	photo	8	0
1999	Pozzulo	46	11-15	.45	30	.13	sim	photo	6	0
2001	Parker	2	9-10	.67			seq	photo	6	0
		2	9-10	.67			seq	photo	6	0
		3	9-10	.50			seq	photo	6	0
		4	9-10	.33			seq	photo	6	0
2001	Lindsay	20	5-6	.65			sim	photo	6	0
		20	7-8	.48			sim	photo	6	0
2002	Memon	10	7-8	.70			sim	photo	6	1
2003	Pozzulo	45	11-15	.45	30	.13	sim	video	6	0

		57	11-15	.17	28	.07	sim	live	6	0
2004	Freire	21	3-4	.58			sim	photo	6	0
		35	5-6	.55			sim	photo	6	0
		33	5-6	.47			sim	photo	6	0
		35	7-8	.41			sim	photo	6	0
		22	7-8	.23			sim	photo	6	0
		10	9-10	.26			sim	photo	6	0
		21	9-10	.36			sim	photo	6	0
		22	11-15	.29			sim	photo	6	0
		14	11-15	.31			sim	photo	6	0
		15	11-15	.11			sim	photo	6	0
2004	Weber	309	11-15	.67			sim	photo	8	0
2006	Pozzulo	46	11-15	.46	30	.13	sim	photo	6	0
		21	11-15	.76	30	.50	sim	photo	6	0

Note. NoCh – total number of children in test. ChildAge – age of child participants. ChildFA – proportion of false alarms for children. NoAdult – total number of adults in test. AdultFA – proportion of false alarms for adults. LineupType: sim – simultaneous; seq – sequential; sho – showup. Persons – number of persons in the lineup. Don't – the possibility to say 'don't know'. Delay – time delay (in days).

The mode of target presentation

The mode of target presentation did not have effect on children's false alarms (live presentation .52; photospread .54; video presentation .45). For false alarms there was no children versus adult differential effect of the mode of target presentation. However, as there was only one study involving adults (as well as children) using video presentation, this finding should be taken cautiously.

2.4.3. Comparison of correct identifications and false alarms

The effects of age (children versus adult) on the proportion of hits in target-present lineups and false alarms in target-absent lineups were examined. There was a significant interaction $\chi^2(1) = 13.26, p < .001, n = 15$, Cohen's $d = 5.40$, $\Phi = .94$, in which the difference between child and adult hit rates (i.e. 57 % and 56%) was smaller than the difference between child and adult false alarm rates (i.e. 53% and 26%). The FSN (267) was acceptably tolerant for future null results.

The interaction was larger for sequential lineups $t(1) = 4.86, p < .04, n = 3, r = .94$ and for showups $t(1) = 4.76, p < .04, n = 3, r = .94$ than for simultaneous lineups $t(1) = 2.99, p < .017, n = 9, r = .71$. Namely, for sequential lineups and showups, the proportion of hits for children was larger than for adults (i.e. 47 % and 38% for sequential lineups and 79% and 54% for showups), but children also made more false alarms than adults (65% vs 18% in sequential lineups and 40% vs 15% in showups). For simultaneous lineups, the proportion of hits was equal (children 57% vs adults 59%). However, adults made fewer false alarms than children in target-absent lineups (children 53% vs adults 31%).

2.5. DISCUSSION

The results of this review indicate that differences between adults and children (all ages combined) appear to have decreased in comparison to previous reviews in terms of overall rates of correct identifications and false alarms. Nevertheless, younger children made fewer correct identifications and more false alarms than older children and adolescents. The fewer members there are in the target-present lineup, the more accurate are children's identifications (especially for showups).

However, with less members in a lineup, the possibility to choose a person by chance (whether correctly or not) increases.

It should be noted that a random effects model was used in this study which was stricter than a fixed effects model (used in Pozzulo and Lindsay's (1998) meta-analysis). This could explain why several findings are different in this review compared to the findings of their meta-analysis (e.g. differences between adults and children in correct identification). However, as the number of hypotheses was larger in this meta-analysis then the results can be more trustful according to FSN statistics compared to Pozzulo and Lindsay's meta-analysis.

2.5.1. Age effects

Kindergarten children's correct identification rate was somewhat lower than that of younger and older children (although not statistically significantly). This is consistent with previous findings (Pozzulo & Lindsay, 1998; Shapiro & Penrod, 1986). Surprisingly, preschoolers' correct identification rate in this study was comparable to adolescents' (whose identification rate was the highest among the child age groups). Pozzulo and Lindsay (1998) and Dekle et al. (1996) found differences between adults and children in correct identifications. However, in the present review no such difference between adults and children (all ages combined) emerged (as also reported by Parker & Ryan, 1993; Gross & Hayne, 1996; King & Yuille, 1986).

Regarding false alarms on TA lineups, there were no effects of child age on false alarms, although children made more false alarms than adults. In Pozzulo and Lindsay's meta-analysis (1998) there were age differences in false alarms between both older children and adolescents compared to adults. However, for younger children this effect was not present. Some previous studies have found false alarms to decrease with age (Chance et al., 1982; King & Yuille, 1986; Davies et al., 1988). Clearly, for TA lineups more research is needed on (i) why children choose and (ii) assisting children to perform better. Also, there exist only a few studies with children which have analyzed this notion more closely.

2.5.2. The number of members in a lineup

A novel finding of the current meta-analysis was that for TP lineup children's identifications were more correct when there were fewer members in the lineup (perhaps partly because of the increase in correct performance that could be caused by guessing). There was no such effect for TA lineups (in which guessing does not contribute to accurate responding).

2.5.3. The effect of delay

Another novel finding is that delay did not have an effect on children's correct identifications. Also, there was no effect of time delay on children's false alarm rates for TA lineups, although Davies (1996) found that delay had stronger effect on children's errors in TA lineups. However, Goodman et al. (1991) and Peters (1987, 1991) found no such effect of delay. Thus, it can be concluded that investigators should not be too concerned about the effect of delay on children's identifications up to the delay intervals employed in the studies reviewed here.

2.5.4. The type of lineup

The type of lineup influenced children's correct identifications, namely for showups the correct identification rate was higher than for simultaneous or sequential lineups (though this was not the case for adults). For TP showups children achieved significantly higher correct identification rates than adults (perhaps due to their propensity to choose), whereas the correct identification rates for simultaneous and sequential lineup were not different between adults and children. However, the number of studies employed showups was small; therefore the conclusions regarding the usage of showups should be taken with care.

For TA lineups, the type of lineup had an effect on children's false alarms in that they made fewer false alarms with showups than with simultaneous or sequential lineups. However, in the children versus adults comparison an age effect was present, namely children made more false alarms than adults for all types of lineups which is similar to the finding of Pozzulo and Lindasy (1998). Steblay et al. (2003) and Dekle et al. (1996) also noted that TA showups produce a higher level of

correct rejections in adults than in children. The results of this review support previous findings which have found that sequential lineups do not increase children's correct rejections (Lindsay et al., 1997; Lindsay & Wells, 1985).

2.5.5. The mode of target presentation

There were no effects of the mode of target presentation (i.e. whether target was presented either live, via video or via photographs) for children, and no difference in performance between children and adults for both TP and TA lineups. Lindsay and Harvie (1998) found similar results to this for TP lineups, whereas they found that for TA lineups correct rejection rates were higher for the slide and video conditions. Lindsay and Pozzulo (1998) found that children correctly rejected TA lineups with a live target presentation less often than adults did. Nine to ten-years-old children in Lindsay and Pozzulo's study were also less likely to make correct rejections from photographs than adults. In conclusion, if a lineup is conducted properly the mode of target presentation will probably not influence significantly the outcome.

2.5.6. Comparison of correct identification and false alarm rates

The differences between child and adult correct identification rates in TP lineups were smaller than the between child and adult false alarm rates in TA lineups (especially for sequential lineups and showups). Adults made fewer false alarms on TA lineups than did children (all types of lineups). However, the differences between children and adults were less present in TP and TA lineups for simultaneous than for sequential lineups or showups (which confirms previous finding in this meta-analysis).

2.6. CONCLUSION

The results of this review indicate that compared to previous reviews the differences between children and adults in correct identifications seem to be diminishing, especially for TP lineups. However, children's performance in TA lineups remains poor. It might be that correct identifications and rejections are driven by different processes, namely correct identification by cognitive memory processes and correct rejection by social as well as cognitive factors (Wells & Luus, 1990).

Although the results of this review indicate that children's performance is better with showups compared to either simultaneous or sequential lineups, this finding should be interpreted with care. More research is needed in this topic because with a showup, the probability of choosing the wrong person (if it is TA) is higher than from usual six-person simultaneous or sequential lineup.

There were no effects of delay. However, it might be that time delays in this review were too short (a maximum of 30 days between seeing person and recognition). Further research is therefore needed in identification employing longer delays. This chapter has examined more closely children's identification performance. In the next chapter children's recall ability is investigated when answering police officers' questions.

3. CHAPTER 3: THE AMOUNT OF JUDICIALLY SIGNIFICANT DETAILS IN ESTONIAN INVESTIGATORS' QUESTIONING STYLES WITH CHILD WITNESSES

3.1. ABSTRACT

Purpose. Interviewers tend to ask many specific and closed questions during the interviews with children. The purpose of the present study was to examine the questioning styles employed in Estonian police interviews of child witnesses, focusing on interviewers' proportionate usage of different types of questions and on the amount of judicially significant information in children's answers to questions.

Methods. Real-life videotaped interviews with 17 children (mean age 8 years 10 months, range 5 to 13 years) were analyzed.

Results. A pattern of long interviewer questions and short child answers was often apparent. More judicially significant details emerged in response to general and cued invitations than to option-posing and suggestive questions. During the interview, the proportion of direct questions was found to decrease over time and the proportion of option-posing and suggestive questions to increase. Longer answers were provided in response to general or cued invitations, whereas option-posing or closed questions produced less information.

Conclusions. Investigative interviews in Estonia rely heavily on directive and option-posing utterances. Therefore, training is strongly recommended to increase (i) the proportion of invitations in interviewers' utterances and (ii) to aid children to recall more judicially significant information.

3.2. INTRODUCTION

In both laboratory and field settings, open memory prompts (e.g. 'Tell me what happened?') tend to elicit longer and more accurate responses than do closed prompts (e.g. 'Did he wear a coat or a jacket?') (Hershkowitz, Lamb, Sternberg, & Esplin, 1997; Lamb & Fauchier, 2001; Lamb, Hershkowitz, Sternberg, Esplin, et al., 1996; Orbach & Lamb, 1999; Price & Goodman, 1990; Wood et al., 1998). However, because children's free recall can be poor, it is important how interviewers question them. It is common for interviewers to misunderstand children's speech or to overestimate their linguistic capacities (Lamb, Sternberg, Orbach, Hershkowitz, & Esplin, 1999). Therefore, it is important to know about appropriate/inappropriate questioning methods and their likely impact on the quality and quantity of information, especially in countries that have not been the focus of previous research.

Only a few studies have actually been conducted concerning on which type of details (contextual or judicially significant) emerge in response to which type of questions (Korkman, Santtila, Westeraker, & Sandnabba, 2008). Also, to date no research on investigators' interviewing styles has been made in Estonia, therefore this study examines Estonian interviewers' questioning styles with child witnesses.

3.2.1. Quality of interviews

Despite the good quality of information elicited by open-ended questions, focused and closed questioning has been found to be dominant in forensic interviews of children in Israel, Sweden, Finland, the United States, and the United Kingdom (Cederborg, Orbach, Sternberg, & Lamb, 2000; Craig, Scheibe, Kircher, Raskin, & Dodd, 1999; Davies, Westcott, & Horan, 2000; Lamb, Hershkowitz, Sternberg, Boat, & Everson, 1996; Sternberg et al., 1996; Sternberg, Lamb, Davies, & Westcott, 2001; Stockdale, 1996; Walker & Hunt, 1998; Santtila, Korkela, & Häkkänen, 2004; Korkman, Santtila, & Sandnabba, 2006; Korkman, Santtila, Drzewiecki, & Sandnabba, 2008; Korkman, et al., 2008). Reliance on open-ended questions during the early stages of interviews (even before the interviewer introduces any information) may not only produce greater amounts of

uncontaminated information but also reduce acquiescence to misleading information (unwittingly) introduced later in the interview (Warren & Lane, 1995). Experts have recommended that forensic interviewers should employ as much as possible such open-ended prompts (Home Office, 2002; Bull, 1992, 1995; Lamb, Sternberg, & Esplin, 1998; Lamb et al., 1999; Home Office, 1992; Milne & Bull, 1999; Saywitz & Goodman, 1996).

The adverse effects of option-posing (giving the child two choices for answering, e.g. 'Did he have a beard or not?') and suggestive interviewing techniques are likely to be strongest when they occur early in the interview and when the children are very young (Orbach & Lamb, 2000). Saywitz and Goodman (1996) demonstrated the negative effects of option-posing, misleading, and suggestive questions when introduced prior to open-ended utterances, and similar effects on the accuracy of free recall have been described by other researchers (Goodman & Aman, 1990; Memon, Holley, Wark, Bull, & Köhnken, 1996; Tobey & Goodman, 1992).

When focused questions are needed to explore forensically important information, experts suggest that these be delayed until as late as possible in the interview (Home Office, 2002; Home Office, 1992; Poole & Lamb, 1998). However, in practice option-posing and suggestive questions are often introduced early in interviews (sometimes as the very first interviewer utterance) and can contaminate subsequent information (Cederborg et al., 2000). Also, it has been found that interviewers tend throughout interviews to ask few open questions and ask many specific and closed questions (Sternberg, et al., 2001; Davies, Westcott, & Horan, 2000; Lamb, Orbach, Sternberg, et al., 2002; Westcott & Kynan, 2006).

In forensic contexts, free recall prompts such as general or cued invitations produce three to five times more forensically relevant information than do focused prompts (i.e. focusing on a specific event, 'Tell me more about what happened that night?') (Lamb, Hershkowitz, Sternberg, Esplin, et al., 1996; Sternberg et al., 1996; Sternberg, et al., 2001).

A recent study of young-in-service police officers' perceptions of their interviewing practices in the UK (Dando, Wilcock, & Milne, 2008) found that such officers often

used techniques which they perceived to be more effective such as establishing rapport, explaining the interview process, and telling the witness to report everything. However, they also reported feeling inadequately trained, under pressure, and generally ill-equipped in their knowledge of how to conduct an effective interview.

3.2.2. Verbal ability

Young children's reports of their experiences may not reflect what they actually remember, perhaps because they cannot meet the cognitive and behavioral demands of the typical interview situation. Restricted language skills may also influence encoding. Bishop and Donlan (2005) found that children's poor recall could be accounted for by their poor initial encoding of the story content (for children with specific language impairment even if the story is presented non-verbally). They suggest a ´syntax first` hypothesis according to which the ability to encode depends on ability to use and understand complex clauses.

Age differences in children's verbal ability may also influence their recall. Greenhoot, Ornstein, Gordon, and Baker-Ward (1999) state that young children's recall may be limited by developmental factors such as their less developed linguistic, narrative, and information processing skills. In addition to their limited ability to describe past events, children's language skills may also affect their understanding of an interviewer's questions. Recently, children with higher verbal intelligence have been found to recall more information spontaneously (Chae & Ceci, 2005), also children who provided more accurate cued recall were less suggestible. Therefore, it would be beneficial for the police officer to evaluate child's verbal skills before interviewing him/her. However, Greenhoot et al. (1999) found that there were no significant relationships between children's language scores and their recall performance of details of their paediatric check-ups.

It has been found that girls are usually better than boys in their language development and verbal facility (Jensen, 1998). Hyde and Linn (1988) found in their meta-analysis that through the preschool and early school years, girls exceeded boys in most aspects of verbal performance (e.g. in vocabulary), although they

claimed that such gender differences may be diminishing (across the decades in some countries). However, it has to be noted that some gender differences in verbal ability may still exist. Leaper and Smith (2004) found that girls were more talkative and used more affiliative speech than did boys, whereas boys used more assertive speech. Still, these differences were relatively small in magnitude. Importantly, Kask, et al., (2007) found that children's verbal abilities were related to their free recall of the target person.

To sum up, conflicting results have been found concerning the influence of children's verbal ability on their free recall, although language does seem to be an important factor influencing children's testimonies.

3.2.3. Children's responses

Young children's free recall can be no less accurate than those of older children if they are interviewed appropriately (Flin, Boon, Bull, & Knox, 1992; Marin et al., 1979). However, younger children tend to remember less information and to provide briefer accounts of their experiences than do older children (Baker-Ward et al., 1993; Lamb, Hershkowitz, Sternberg, Esplin, et al., 1996; Ornstein, Gordon & Larus, 1992; Sternberg et al., 1996; Marin et al., 1979; Davies & Flin, 1988; Saywitz & Goodman, 1996; Chae & Ceci, 2005; Kask, et al., 2007) or adults (Marin et al., 1979; Dent & Stephenson, 1979; Davies, Tarrant, & Flin, 1989).

When children are trained to respond to open-ended prompts in forensic contexts, they produce more information when prompted to 'tell everything' about the alleged abuse (Sternberg et al., 1997). Option-posing, 'yes/no', and suggestive interviewer utterances are more likely to elicit inaccurate information because they implicitly encourage children to acquiesce to the suggested information or to guess (Dent & Stephenson, 1979; Poole & White, 1993; Ceci & Bruck, 1995; Poole & Lindsay, 1998).

Korkman, et al. (2008) in Finland studied the sequence of the type of details (contextual or judicially significant) in interviews with children and found that almost 80% of details containing judicially significant material produced by the

children was followed by either directive or option-posing question (and not by facilitators or invitations). This indicates that interviewers continue to rely on those types of questions even after the child had provided significant information (and thus demonstrated relevant ability) and therefore failed to follow-up information provided by the child in an adequate way. However, as the number of contextual and judicially significant details provided in children's responses has not yet been extensively researched, this notion is more closely examined in the present study.

There appear to be no studies which have examined where during the course of the interviews which type of questions are more often used. Therefore, the present study examines this notion more closely.

3.2.4. Present study

The present study examined the questioning styles in Estonian police interviews with children, and the information provided by children in those interviews. Estonian police officers have received training about questioning suspects and witnesses during their general studies, however, they have not had any specific training in questioning children. It is hypothesized that (i) interviewers will use more option-posing and direct questions than prompting of children's free recall (e.g. 'Tell me what happened?'), (ii) children's answers will be longer in response to free recall requests than to questions, (iii) the proportions of usage of different questioning techniques will vary across time within interviews (i.e. free recall is used more in the beginning of an interview, whereas questions are used more near the end of an interview), and finally (iv) cued and general invitations will produce more contextual and judicially details than option-posing questions.

3.3. METHOD

3.3.1. Participants

Videotaped interviews with 17 children were provided from two police units in Estonia. Of the interviewees eight (47%) were male and nine (53%) female whose mean age was eight years and ten months (range 5 to 13 years). The interviews had been conducted in 2004/2005. Mean length of the interview was 23 minutes (range

8 to 55 minutes). The study was conducted only with 17 interviews because it was hard to get access to the police files of those interviews due to the law and procedures in Estonia. The author tried to get access to a larger sample of interviews but this was not possible because of lack of interest or fear of evaluation (for example, from one police station only those few interviews were made accessible to me which were conducted by the police officers *not* working in that department any more). Therefore, only 17 interviews were involved in this study.

Fourteen (82%) of the interviewers were female police officers and three (18%) were males. As only three interviewers were males, it was not possible to conduct any analysis involving the gender of the interviewer. (No information was available concerning the officers' experience, the outcome of the interviews, or what proportion led to trial or conviction.)

3.3.2. Coding

Videotapes of the interviews were transcribed by police officers themselves. Two trained raters reviewed the transcripts and categorized each interviewer utterance, defined by a 'turn' in the discourse or conversation (Orbach & Lamb, 2000). The categories by Lamb and colleagues (Lamb, Hershkowitz, Sternberg, Esplin et al., 1996; Lamb, Hershkowitz, Sternberg, Boat, & Everson, 1996) and by Yuille and Cutshall (1986) were used to characterize the interviewer utterances and to measure the amount of new information provided by the children in each response by tabulating the number of details, operationally defined as the smallest units of interview relevant information. Details involved the mentioning of individuals, objects, events and descriptions of different features (e.g. appearance, actions, and locations). Details were only counted when they added to the understanding of the target incidents, so restatements of facts were not counted.

The categories were as follows (see Aldridge et al., 2004, for fuller overview):

1. *Invitations* consist of ways of prompting free-recall responses from the children and contain two types of invitation, (i) <u>general invitations</u> (e.g., 'Tell me everything that happened') and (ii) <u>cued invitations</u> (where

reference is made to detail(s) mentioned earlier by the child, e.g., 'You mentioned that he punched you. Tell me everything about the punching').

2. *Directive utterances* attempt to redirect the child's attention on details which have been already mentioned (e.g., 'When did it happen to you?' when the child had disclosed that something happened or 'What was he wearing?' when the child had mentioned a man).

3. *Option-posing utterances* try to focus the child's attention on details that the child had not previously mentioned. These utterances prompt the child to affirm, negate, or select an investigator-given option but do not imply that a particular response is expected (e.g., the investigator might ask, 'Did he tell you to touch him over or under your clothes?' when the child had mentioned being touched).

4. *Suggestive utterances* are questions stated in such a way where the interviewer communicates what response is expected (e.g., 'He forced you to do that, didn't he?') or assume details that have not been revealed by the child (e.g., child: 'We laid on the bed'; interviewer: 'He laid on you or you laid on him?').

5. *Verbal affirmations* are interviewer's responses to children's answers (e.g. 'Yes, I see' etc.).

6. *Confirming comments* are interviewer's summaries what the child has said (e.g. 'As I understand... ' etc.).

7. *Explanations* are interviewer utterances not related to the alleged incident under investigation (e.g., references to the interviewer's and child's roles).

8. *Praise* is interviewer utterances which encourage children to continue with their testimony (e.g. 'Good! Now I want to talk you about...' etc.).

3.3.3. Details reported by the child

The number of words in children's responses was counted, and the children's responses were coded for the types of details they included, based on a coding scheme developed by Korkman, et al. (2008) as described below:

1. *Judicially significant details* concerning the crime or details clearly pointing to the non-existence of evidence for a crime.

2. *Contextual details* which added to the understanding of the child's situation and to circumstances surrounding the investigated event (but not of judicial significance).

Details (both contextual or judicial), as stated above, were coded as significant only when the child had introduced them into the interview and only the details that added new information to the understanding of the events were counted.

It was decided to divide each child's interview into four equal quarters regardless of how many questions were asked (for example, in an interview where 300 questions were asked, the quarters consisted of 75 questions whereas when 120 questions were asked then quarters consisted of 30 questions).

Thus, the analysis of the interviews involved a quantitative analysis of the utterances with which the investigator tried to elicit information and analysis of the children's responses, assessing the number of informative details about the reported incident.

Two independent raters coded randomly 30% of the transcripts to measure the inter-rater reliability. The agreement for the interviewer utterances was 93% (Cohen's $\kappa = .92, p < .001$) and for the details provided by the children 89% (Cohen's $\kappa = .86, p < .001$). Disagreements between raters were solved through discussing the particular cases.

3.4. RESULTS

First, information concerning the proportion of different question types will be presented. Then will be examined the number and the proportion of information units in the children's answers. Finally, the type of information reported by the children is reported.

3.4.1. The proportion of questions in interviews

The interviewers on average asked more than one question per utterance (M = 1.50, SD = 0.70, range 1 to 9). As hypothesized, option-posing and direct questions were used mostly by the interviewers $\chi^2(8) = 1178.30$, $p < .001$, (see Table 3.1), followed by confirming comments, suggestive questions, and verbal affirmations. The proportion of either general or cued invitations was relatively small (see Table 3.2 for example quotations from the interviews).

Table 3.1. *The Proportion of Questions in Interviews.*

Category	%
Explanation	3.39
General invitation	0.55
Cued invitation	1.89
Verbal affirmation	4.55
Confirming comments	13.34
Direct questions	30.44
Option-posing questions	38.90
Suggestive questions	5.53
Praise	1.07

As stated above, the interviews were then each divided into four equal quarters and the proportion of questions in the four quarters was examined. There were significant differences between quarters for explanations $\chi^2(1, N=17) = 15.14$, $p < .001$, general invitations $\chi^2(1, N=17) = 4.22$, $p < .05$, and praise $\chi^2(1, N=17) = 5.64$, $p < .02$ (see Table 3.3). There was an almost significant effect for cued invitations $\chi^2(1, N=17) = 3.40$, $p = .065$. Namely, explanation and praise were more often used

in the first quarter whereas general (and cued) invitations were more often used in the second and third quarter.

Table 3.2. *Quotations from the Interviews.*

General invitation	*What happened then? - I went to the elevator and then he came along. The man pushed the button, turned around and didn't let me out.* (Interview with 8-year-old girl)
Cued invitation	*Let's talk more about this man. Do you remember any other things concerning him? - He was tall but not so tall as my Mum is.* (Interview with 8-year-old girl)
Verbal affirmation	*His hair was red... - His hair was red? I see.* (Interview with 10-year-old boy)
Confirming comment	*I did not see, he had something covering his face – Covering his face? – Yes.* (Interview with 10-year-old boy)
Direct question	*How many times did he punch you? - Three... three times.* (Interview with 7-year-old girl)
Option-posing question	*So you said you were both on the bed. Did you sit there or did you lie down? - Lie down.* (Interview with 8-year-old boy)
Suggestive question	*All the time? How come all the time, was it every...every day? - Almost every day.* (Interview with 7-year-old girl)
Praise	*Do you remember anything else about this? – No. – Well done. Then we go on.* (Interview with 6-year-old girl)

Table 3.3. *The Proportion of Questions (in percentages) in Different Quarters of Interviews.*

Category	First quarter	Second quarter	Third quarter	Fourth quarter
Explanation	7.33	1.34	2.44	2.44
General invitation	0.49	0.86	0.73	0.12
Cued invitation	1.34	1.71	3.06	1.46
Verbal affirmation	3.42	4.77	3.91	6.10
Confirming comments	13.08	14.67	13.20	13.66
Direct questions	33.13	34.11	29.95	24.63
Option-posing questions	34.72	36.06	40.71	44.15
Suggestive questions	4.03	6.11	5.26	6.71
Praise	2.44	0.37	0.73	0.73

3.4.2. Children's responses to the questions.

The number of information units provided by the children in response to each question was counted. When several questions were asked in an utterance (which happened in most of the cases), the last question asked was coded as the ´stimulus` question. The influence of the different types of questions on the total amount of information contained in the children's answers was analyzed with a one-way ANOVA which indicated a significant effect $F(1,8) = 11.76$, $p < .001$, $\eta^2 = .042$, (see Table 3.4). Scheffé post-hoc analysis revealed that longer answers were provided in response to (i) cued invitations compared to confirming comments, direct, option-posing, and suggestive questions ($p < .001$) and (ii) direct questions compared to option-posing questions ($p < .001$).

However, most interviewer utterances ended with option-posing or closed questions which produced less information. Again, relatively small proportions of either general or cued invitations were used as the ´stimulus` question.

Table 3.4. *The Number and the Proportion of Information Units in Response to Different Categories.*

Category	Mean	SD	%
Explanation	3.24	8.09	1.1
General invitation	8.38	24.30	0.7
Cued invitation	9.63	14.31	2.1
Verbal affirmation	9.88	9.03	0.4
Confirming comments	3.55	4.90	3.5
Direct questions	4.94	8.48	38.5
Option-posing questions	2.66	4.81	47.2
Suggestive questions	2.79	4.49	6.5
Praise	2.00	0.00	0.1
Total mean	3.80	7.21	100.00

Note. Mean – mean number of information units in response to questions; % - proportion of the last 'question' in the end of utterances in percentages.

3.4.3. Type of information reported by the children

Table 3.5 demonstrates the influence of different types of questions on the amounts of detail reported by children. (The final column in Table 3.5 contains the percentages from each of the different types of utterances.) A one-way ANOVA found a significant effect of questions type for contextual details $F(1,8) = 3.67$, $p < .001$, $\eta^2 = .032$ and another one-way ANOVA for judicially significant details $F(1,6) = 15.01, p < .001$, $\eta^2 = .093$. Scheffé post-hoc analysis indicated that children reported significantly more judicially significant details in response to (i) general invitations compared to all other question types, (ii) cued incitations compared to option-posing and suggestive questions, and (iii) direct questions compared to option-posing questions. No post-hoc effects were found for contextual details.

Table 3.5. *The Number of Contextual and Judicially Significant Details Reported in Response to Different Categories of Interviewer Utterances.*

Category	Contextual		Judicial		%
	M	SD	M	SD	
Explanation	4.47	9.36	0.00	0.00	3
General invitation	3.00	2.27	36.67	54.04	1
Cued invitation	10.60	15.14	10.97	14.61	2
Verbal affirmation	8.40	7.83	11.75	11.35	5
Confirming comments	5.00	5.50	3.41	4.02	13
Direct questions	4.59	7.54	6.90	9.84	30
Option-posing questions	3.07	4.09	3.62	5.99	39
Suggestive questions	2.93	4.26	3.47	4.95	6
Praise	2.00	0.00	0.00	0.00	1

Note. % - the proportion of this type of question in interviews in percentages.

3.4.4. An international comparison of the results

The results from the present study were compared to those of a number of international studies in other countries (Cederborg, 2000; Lamb et al., 2000; Orbach & Lamb, 2000; Sternberg et al., 2001, Korkman et al., 2006, results presented in Table 3.6). (The interviews analyzed in the study from England and Wales were conducted by interviewers who were all guided by the Government's 'Memorandum of Good Practice', which states how forensic child interviews should be conducted.) As can be seen, the general structure of the interviews analyzed in the present study shares many characteristics with other studies. As the sample size of the present study was small, it could be that the results are not very representative of investigative interviews in Estonia. However, all studies in this comparison share the fact that invitations were in the minority among the interviewer utterances, while there was a clear predominance of directive and option-posing utterances. In comparison to the other studies, the present study there were more option-posing questions and less invitations and verbal affirmations, while the percentages of directive and suggestive questions was quite similar to those of the other studies.

Table 3.6. *The Frequency of Different Interviewer Utterance Types in Percentages: An International Comparison.*

Utterance type	Israel n=50	USA n=145	Sweden n=72	England n=119	Finland n=12	Estonia n=17
Invitation	5	3	6	6	4	3
Verbal affirmation	-	17	-	12	13	5
Direct questions	52	31	41	47	22	30
Option-posing questions	33	31	39	29	30	39
Suggestive questions	10	5	14	5	23	6

Notes. Utterance types are presented in percentages (%).

3.5. DISCUSSION

This study examined the questioning styles employed in Estonian police interviews of child witnesses. It also focused on the amount of judicially significant details in children's answers to questions. Option-posing and direct questions were used mostly by the interviewers. Children's answers were longer in response to free recall requests than to questions. The proportions of usage of different questioning techniques varied across time within interviews, namely explanation and praise were more often used in the first quarter of the interview whereas general (and cued) invitations were more often used in the second and third quarter. As hypothesized, children reported more judicially significant details in response to invitations than to option-posing and suggestive questions. The results of this research demonstrate that Estonian Police officers' interviewing skills and problems in questioning children are comparable with those in other countries. When interpreting the results in the present study it should be noted that the sample is small (obtaining access to such sensitive material is a challenge for 'outsiders') and may not be representative.

Similarly to Korkman et al. (2008), a pattern of multiple questions per utterance (on average one and a half questions at a time) and short child answers was clearly apparent. A large proportion of the interviewers started to use specific questions

early in the interviews and ended the free narrative phase (too) quickly. This may be because the interviewers were aware of evidential requirements for specific details, but doing this fragments the child's account and increases the opportunity for errors in answering, thus having a potential effect on the perceived credibility of the child's interview (Westcott & Kynan, 2004). The continuing failure to encourage a free narrative (or to try and maintain it once the child had begun) with many of the children is a concern, since this phase allows them to present their 'story' in their own words (Westcott & Kynan, 2006). It has to be noted that every retrieval attempt generates a new memory trace, which may affect retrieval of some aspects of the original memory. This could result in more likely recalling details which were recalled previously and less likely to recall details which were not recalled previously (Valentine, 2007).

In the beginning of the interview the interviewers used more explanations (i.e. what is going to happen, what is expected from child etc.) and praised the child in order to build rapport. General (and cued) invitations were more used in the middle of the interview. The proportion of direct questions decreased with time, and the proportion of option-posing and suggestive questions increased near the end of the interview. These changes in interview dynamics contradict those of Cederborg et al. (2000) who found in Sweden that option-posing and suggestive questions were often introduced early in interviews.

One possible explanation for the overall large number of option-posing questions may be that officers are not aware of the negative effects of such closed questioning. Also, they might feel pressured by legal demands associated with investigative interviews (e.g. if they do not get a 'proper' statement from the child the culprit may not be prosecuted).

Indeed, it is demanding to keep a child's focus on a not very pleasant topic which may lead to confusing and overlong statements from the interviewer (e.g., 'I will explain you now what is going to happen. This thing in the corner is a video camera which will record our conversation. The purpose of recording our conversation is that it saves time, we can get it over with more quickly, and I can ask you things which I need to ask you. And I say also that the tape is not shown to any stranger, it

goes safely to court archives and is only seen by people who have to see it. Allright. And about the interview I want to tell you that if you have to go to toilet during our conversation you can do it or if you have questions, you can ask them freely. If you don't understand my questions, then say it right away. Then I will rephrase my question. And also I'd note that if you don't remember or don't know the answer, then say so. This is also an answer if you say you don't remember or don't know. You don't have to say ´yes´ right away. And if you need to take time out then also say that to me. All right. How do you feel right now?`). It is not surprising (taking the results of this study into account) that the child responded to such a long interviewer utterance with a short answer.

As hypothesized, longer answers were provided in response to general or cued invitations. Option-posing or direct questions produced less information, which supports previous findings (Lamb, Hershkowitz, Sternberg, Esplin, et al., 1996; Sternberg et al., 1996; Sternberg, et al., 2001; Dent & Stephenson, 1979; Hutcheson et al., 1995). Korkman, Santtila, and Sandnabba (2006) also found that directive and option-posing questions were mainly used, and invitations were associated with the longest answers from children and were (together with directive utterances) the type of interviewer utterance that elicited most of the new details from the children. It has been suggested that focused questions should be asked as late as possible in the interview (Home Office, 2002; Poole & Lamb, 1998).

Although invitations (and open-ended questions) produce longer statements than do option-posing questions, we do not know for ´real-life` interviews whether those details are correct or not. However, in studies where details of the event are known to the researchers, open-ended questions produce more accurate answers than do closed or suggestive questions (Korkman et al., 2006, for a review).

3.5.1. Type of information reported by the children

Different types of questions influenced the amount of both contextual and judicially significant details reported by children. As hypothesized, children reported more contextual and judicially significant details in response to invitations than to option-posing and suggestive questions. Korkman et al. (2008) indicated that interviewers'

seem to believe that focused questions can be the most efficient in eliciting information from the child because they believe children will not otherwise talk about the events under investigation. However, their study demonstrated that even when the children provided new judicially significant information, 80% of the subsequent questions were directive, option-posing or suggestive.

3.5.2. An international comparison of the results

When comparing the results of the present study with those of studies in other countries, it should be noted that even the categories directly corresponding to the ones used by Lamb and colleagues might have been coded differently in some of these studies. However, the general characteristics of the main categories may be clear enough for a comparative discussion.

When taking into consideration the frequency of the different interviewer utterance types, it is clear that investigative interviews in Estonian sample share the heavy reliance on directive and option-posing utterances found in the interviews conducted in other countries such as Sweden (Cederborg et al., 2000), Finland (Korkman et al., 2006), Israel (Orbach et al., 2000), England and Wales (Sternberg et al., 2001), and the USA (Lamb, Sternberg, & Esplin, 2000). However, research using a larger sample of Estonian interviews is probably needed before drawing conclusions on the general quality of interviews conducted in that country.

3.5.3. Suggestions for the future

Agnew, Powell, and Snow (2006) pointed out that the quality of evidence obtained from children is likely to be very dependent on the degree to which police interviewers adhere to best-practice guidelines, as well as the children's general experience with an-open-ended style of communication.

Dando, et al. (2008) report that over 60% of police officers noted that their interview training had been far too suspect-biased, with over 70% reporting that they felt ´not very well` or ´not at all well` equipped to interview witnesses. Other officers have reported ´usually` or ´almost always` feeling pressured to complete

witness interviews quicker than they would like (Kebbell, Milne, & Wagstaff, 1999; Clarke & Milne, 2001). Davies, Wilson, Mitchell, and Milsom's (1995) evaluation of field interviews by police in the UK noted that in 43% of interviews the child was needlessly rushed from the free recall to the questioning phase.

Similarly, Wright, Powell, and Ridge (2007) studied police officers' perceptions regarding their performance in a mock interview with a five to seven-year-old children and found that officers' perceptions of what constitutes an effective interview were not entirely consistent with those held by experts in forensic interviewing. Namely, the majority of interviewers perceived that the locus of control during the interviews rested primarily with the child and/or the environmental setting and not on the skills of the interviewers (in using open-ended prompts and questions).

However, it has been found that when interviewers use a protocol designed to interview children, they adhere to recommended practices more than interviewers who do not use the protocol, and children interviewed using the protocol provide more free-recall details than do children interviewed without the protocol (Orbach, Hershkowitz, Lamb, Sternberg, Esplin, & Horowitz, 2000).

Better police training in the adoption of best-practice guidelines in interviewing children is required. Therefore, more structured methods such as the NICHD protocol (Orbach, et al., 2000), Achieving Best Evidence (Home Office, 2002), or cognitive interview (Fisher & Geiselman, 1992) which has been found to be effective with children (Milne & Bull, 2003; Holliday & Albon, 2004), are strongly recommended to structure the interviews, increasing both quality and quantity. The NICHD protocol emphasizes especially the importance of encouraging free recall and asking 'appropriate' questions which (as the present study demonstrated) are especially important for increasing the number of judicially significant information in children's recollections.

Lamb, Sternberg, Orbach, Hershkowitz, Horowitz, and Esplin (2002) found that intensive training in the use of a highly structured interview protocol followed by

continuing supervision, monthly day-long seminars, and feedback on all field interviews, yielded improvements in police officers' interview quality.

There could be several reasons why training is not very well applied in practice. The officers might not have been interested in training (or the particular training session was not performed at a good level of quality). This could explain why they could not do what they learned during the training sessions. Also, there can be difficulties in changing shift in thinking, i.e. if the police officer is used to do his/her work in a certain way for years it can be hard to start to think differently (Westenholz, 1993). Concerning this notion, applying newly acquired knowledge into practice is time consuming and is easier when the situation is not stressful. However, in stressful situations police officers' questioning styles may go back to old questioning styles because they are more used to and/or self-confident in those techniques.

When applying new techniques into practice, the cognitive load of the new task is relatively large, therefore, it may seem difficult and requires several repetitions to consolidate the knowledge. Cognitive Load Theory (Sweller, 1988) states that optimum learning occurs in humans when the load on working memory is kept to a minimum to best facilitate the changes in long term memory. Sweller found that learning requires a connection to the schematic structures of long term memory. If the schema connection to long term memory is not made, the learner will likely forget the material, and learning will not occur. In order to achieve the schematic change, and move from clumsy to effortless performance, the learner must become more familiar with the material he/she is studying. As a result of this familiarisation, the cognitive processes associated with the material become altered, and the material is handled more efficiently by working memory which, in turn, facilitates the connection to long term memory schemas.

Finally, in ordinary human conversation many focused and closed questions are used repeatedly but in police interviewing the opposite is expected, therefore, it is a real challenge for police officers to change their questioning techniques and style of conversation which differs from everyday-life conversations.

3.6. CONCLUSION

This study found that the questioning styles of Estonian Police officers are similar to those in other countries. More closed than open-ended questions and/or invitations were asked. Therefore, more attention should be addressed to those issues in training police officers how to interview a child. Police officers should avoid contamination of witness memory and use sensitive techniques to obtain as much and as accurate and detailed information as possible. It should be assured that knowledge is integrated into practice in order to enhance the quality of police interviews with children, in Estonia as well as in other countries. It is costly both to children and to society if child abusers are not brought to justice. Therefore, in the next chapter the effects of a different (innovative) method of obtaining information from witnesses are closely examined.

4. CHAPTER 4: TRYING TO IMPROVE YOUNG ADULTS' PERSON DESCRIPTIONS

4.1. ABSTRACT

Purpose. Witnesses' person description performance has repeatedly been found generally to be rather poor. This study examined on the effects of using a person in young adult witnesses' visual field as a comparison (or ´standard`) to assist their recall of a previously seen, different person. Also the effect of each witness' physical similarity to the target in the witnessed event was examined (the ´own-anchor effect`).

Method. The participants were young adults (N=92) who saw a male stranger and who later described and answered questions about that person (half of them using a male or a female ´standard`).

Results. There was no overall effect of the use of a standard but the use of a ´typical` male standard increased the amount of information about characteristics that would otherwise not be remembered well. The own-anchor effect was present only for some characteristics.

Conclusions. The findings reveal that use of a ´standard` does not generally seem to assist witnesses to recall more information about the characteristics about a once-seen person. The methods to increase information about ´difficult` characteristics and especially to increase children's recollections should be examined further as well as the typicality of the standard. Further research should also emphasize studying the possible own-anchor effect in different settings.

4.2. INTRODUCTION

People have difficulties in describing other persons. The amount of available information about a perpetrator is very crucial when a police officer has to make a rapid decision about the perpetrator (especially using solely witnesses' person descriptions). When the descriptions of the culprit's characteristics such as age, height, weight etc. differ a lot from the culprit's actual characteristics, then much police time could be wasted which could result in the culprit's escape. Several attempts have recently been made by various authors to try to improve child and adult witnesses' person descriptions but these have had limited success.

One suggestion aimed specifically at increasing the amount of information children provide about a given stranger is to invite children to make judgments relative to a ´standard`, for example a person in his/her visual field at the time of giving the person description (Sporer, 1996; Meissner, Sporer, & Schooler, 2006) which has proven to be unsuccessful with children (Kask, et al., 2007). However, no previous studies seem to have tested this notion directly on adults, i.e. whether this method aids adult witnesses to recall more information. Since this method requires witnesses to understand the ´representative` nature of the ´standard` its possible efficacy should first be investigated with adults (who are likely to understand the ´representation`) before it is investigated with children (who may have some difficulty understanding the ´representation`).

4.2.1. Why are people poor at describing other people?

Many witness descriptions contain information concerning characteristics such as height, weight, and age of the culprit (Kuehn, 1974; Van Koppen & Lochun, 1997; Yuille & Cutshall, 1986), although such information can be erroneous (Yuille & Cutshall, 1986). Also, general physical characteristics such as height and weight play a role in eyewitness identifications (Wells, 1984), but there has been little systematic investigation of this issue.

It is generally accepted that estimates of height and weight are relatively inaccurate and thus might be influenced by personal determinants (see Clifford & Bull, 1978; Janssen & Horowski, 1980). Witnesses' own characteristics might influence their

person descriptions. Also, the estimations about physical appearance characteristics of perpetrators might also be influenced by witness knowledge about population norms (Meissner, Sporer, & Schooler, 2006) (for example, the average height and weight in the population). Therefore, as people's performance in describing other people is poor compared to correctly recognizing (see Chapter 1), then more efficient methods should be applied to make person descriptions easier for both children and adults.

4.2.2. Attempts to improve (children's) person descriptions

There have been several attempts to improve children's person descriptions which have included the usage of anatomical dolls (Saywitz, Goodman, Nicholas, & Moan, 1991; Bruck, Ceci, & Francocoer, 1995), a body map (Wilcock, Morgan, & Hayne, 2006), or a 'Draw and Tell' technique (Poole & Lamb, 1998).

Anatomical dolls can be considered as a useful adjunct to enable children to demonstrate actions which they are unable to describe verbally (Westcott, Davies, & Clifford, 1989). However, there are concerns that the use of dolls lead to an increase in incorrect as well as correct information about events (Leventhal, Hamilton, Rededal, Tebano-Micci, & Eyster, 1989; Saywitz et al., 1991; Ceci & Bruck, 1995) or may encourage fantasy among younger and more suggestible children (Thierry, Lamb, Orbach, & Pipe, 2005). The use of body maps is often encouraged as a non-suggestive alternative to dolls (Lamb, 1994), but Wilcock, et al. (2006) found they led five to six-years-old children to volunteer as much inaccurate as well as accurate information (see also Aldridge, et al., 2004).

Asking children to draw during investigative interviews sometimes occurs (Poole & Lamb, 1998). It is argued that drawing can help children to recall more details (Butler, Gross, & Hayne, 1995) and also to decrease anxiety in communication with the interviewer (Gross & Hayne, 1998, 1999a, 1999b; Malpass, 1996). Some studies have reported that drawing increased free recall accuracy in 5-10 year old children (Butler et al., 1995; Gross & Hayne, 1998), but others have claimed that children aged 5-6 years show an increase in incorrect as well as correct details when drawing

is used (Bruck, Melnyk, & Ceci, 2000; Strange, Garry, & Sutherland, 2003; Bruck & Melnyk, 2004).

However, as people's persons descriptions and spontaneous recall is usually quite poor, it has been suggested that providing people with possible ranges (to evaluate the person's height and weight) or a colour plate (to evaluate the correct colour of the person's clothes and hair), may lead to better results for some aspects of person descriptions (Sporer, 1996; Meissner, Sporer, & Schooler, 2006).

A way that has been suggested to increase the quantity of information witnesses provide about once-seen stranger, is to invite them to provide descriptions relative to a 'standard', for example a person now in the person's visual field (Sporer, 1996; Meissner, Sporer, & Schooler, 2006). This means that people can provide their recollections about an once-seen person using the person who is questioning them (e.g. police officer) as a standard when recalling different stable (i.e. height, weight, and age) and unstable (i.e. type and colour of hair and clothes) characteristics about the once-seen person.

Kask et al. (2007) found that provision of such a standard did not help children overall to recall more information about a once-seen person, but the younger children's performance when answering some questions was better when the standard was available. However, some children in that study may not have fully understood what was meant by trying to use the standard as an aid. Therefore, the present study sought to investigate if the use of a standard would assist adults. If it did, then it would be worth devoting effort to assisting children to understand it.

Using a standard could particularly assist witnesses to remember more concerning characteristics which are not otherwise remembered well. As stated in Chapter 1, adult witnesses have been found to recall better characteristics such as gender, height, age, face, hair, clothing, and poorer inner face features and figure (Kuehn, 1974; Ellis, 1990; Sporer, 1992b; Lindsay, Martin, & Weber, 1994; Van Koppen & Lochun, 1997; Pozzulo & Warren, 2003).

4.2.3. Gender differences in person descriptions

Characteristics of witnesses, such as gender, have sometimes been found to affect the descriptions they provide of once-seen persons (see Chapter 1 for overview). Kask et al. (2007) found that with the male standard the gender of witness had no effect, but with the female standard boys performed better than girls. One possible explanation is that kindergarten teachers and school teachers in Estonia are usually women. It might be that the children are more used to seeing women and therefore find more useful a standard which is familiar to them.

However, this notion does not explain clearly why the differences did not emerge in case of the girls. It could be that when the gender of the target and standard are the same, transfer effects may occur, for example some characteristics of the standard may also be attributed to the target. This is less likely to happen if the target and standard are of different genders. Therefore, it is worthwhile to study whether any effects of gender of witness and gender of standard are present also in adults.

4.2.4. Cross-ethnic effects

The cross-race effect (also termed 'own-race bias', which means that individuals recognize better 'own-race' faces) has been demonstrated in field settings (Brigham, Maass, Snyder, & Spaulding, 1982; Platz & Hosch, 1988) as well as in numerous laboratory studies (see Meissner & Brigham, 2001, for a review; and Pezdek, Blandon-Gitlin, & Moore, 2003, for a recent study with children). [A more comprehensive overview of the 'cross-race' effect is presented in Study 4 (Chapter 5).]

Fallshore and Schooler (1995) speculated that such an effect may be due to differential reliance on configural vs. featural processing for own vs. other race faces. They stated that there was *no* significant relationship between description accuracy and identification performance for 'other-race' faces. This supports the view that 'other-race' faces may be recognised in a more featural manner than own-race faces (Rhodes, Brake, Taylor, & Tan, 1989; Sporer, 2001a). As the cross-race effect has been shown to have existed for face recognition and person identification,

it could also have an effect when a standard from one ethnicity is used with witnesses of various ethnicities.

4.2.5. Own-anchor effect

Relatively few studies have examined whether the physical characteristics of witnesses can affect the descriptions they provide of targets/perpetrators. For example, Flin and Shepherd's study (1986) used a range of male targets varying in height and weight. Members of the public were asked for directions by one of these targets in a busy city centre. Once the target had disappeared from view, the participants were asked by a second confederate to estimate the target's height and weight. Participants tended to underestimate the height of the target. The greatest mean underestimate was 4.91 inches for the tallest target (78 inches). It is interesting that the target for which there was a mean overestimation of height was the smallest target. A similar situation was found for weight estimates in that the weights of the heaviest targets were consistently underestimated and the lightest targets overestimated.

Also, in their study observers' own height and weight tended to effect their estimations of the perpetrators' height and weight, but this 'own-anchor' effect was found only for male observers describing a male target. Thus, it would appear that one's own physical characteristics might affect judgments about the height and weight of other individuals. Taller participants tended to be more accurate in their height estimates of the targets than did shorter participants as there was a general tendency to underestimate the height of the targets. Inaccurate estimates for the tallest, lightest, and heaviest targets tended toward rather than away from population norms, perhaps indicating a general regression to the population mean. However, there seems to be little theoretical understanding of how such biases might work. As one possibility Manis and Paskewitz (1984) have proposed a judgment model which suggests that prior experience with some members of a given category affects the assessment of other members of the same category (see Chapter 1).

4.2.6. Present study

The aim of the present study is to determine if young adults can use a 'standard' as an aid to help them provide more correct information about a once-seen person. If using a standard is found to be effective with adults, then further adapting how the standard could best be used by children would be worthwhile. The study examines the effects of gender and witness ethnicity. Furthermore, the possible interaction between gender of the standard and gender of the participants is investigated. As characteristics of witnesses, such as gender, have also been found to affect the descriptions they provide of once-seen persons (Yarmey, 1993), possible gender effects on participants' descriptions and their possible relationship with the own-anchor effect is also examined.

4.3. METHOD

4.3.1. Participants

The participants were young adults (N=92) with the average age of 18 years and 8 months (range 17 to 58 years), 22 of them were Caucasian (14 males and 8 females) and 70 of South-Asian origin (23 males and 47 females). The participants were recruited by notices distributed at colleges of Leicester, UK. All participants spoke English as their first language.

4.3.2. Design

A 2 (gender; male vs female) × 2 (presence of a standard; standard vs no standard) × 3 (typicality of standard; typical male, atypical male, typical female) × 2 (ethnicity; Caucasian vs South-Asian) between-participants design was used.

4.3.3. Procedure

A male or a female experimenter entered the room in the college where participants were together in groups with their teacher. The experimenter introduced himself/herself and had a brief conversation with participants. Then a male stranger entered and was introduced as a friend of the experimenter. He said a few words to the participants and gave a map to the experimenter. They had a little talk about

why he was late. Then he claimed that he had to go back to the university and could not stay any longer. They said goodbye to each other and he left. The participants saw the stranger for about one minute.

Later the experimenter asked if the participants would be willing to perform several tasks and then handed response sheets to the participants. First, the experimenter said: ´Please describe my friend who was in the classroom a while ago. What do you remember about him? Write down only what you remember.`.

When the participants had finished writing, the experimenter said: ´Is this all you can remember? It is all right, if you do not remember more. Now I am going to ask you some questions about the stranger you saw. When you answer my questions, please write down only the things that you remember. If you are not sure or if you do not remember, then write ´I don't know` or ´I don't remember`. OK?`.

Then a set series of questions about the once-seen person's figure, height, gender, clothes, hair etc. were asked (see Appendices 1 and 2 for figures and questions used in this study). For half of the participants these ´appearance` questions were asked using the data collector as a ´standard`. For example, without the standard the participant was asked (a) ´How long was his hair?` and (b) ´There are some faces on this paper. Which face did my friend have?`. With standard the questions were (a) ´My hair is this long. How long was his hair?` and (b) ´There are some faces on this paper. I have this kind of face (points). Which face did my friend have?`. Three different people (typical male, typical female and atypical male) were used as a standard (with different groups of participants) to examine the issue of ´If the standard happens to be very different from the target, could that confuse or interfere with witness performance?`.

A week later the participants filled in a questionnaire where the questions about their own stable characteristics such as height, figure, face shape, mouth, nose, and ears were asked.

4.3.4. Coding

The (i) free descriptions and (ii) responses to the appearance questions were coded (i.e. scored) for the total number of semantic units (SUs) (i.e. correct and incorrect, using the same method as Poole & White, 1991). The coding was performed by two different coders who were unaware of the purpose of this study. Intercoder agreement for descriptions and answering questions was 72% and 87%, respectively.

The contents of the free descriptions were divided into the following appearance categories: gender (male or female), age, face, name, body (height and body build), hair (colour and length), clothing (including the colour of the clothes), and boots. The rules for scoring were: one point for correctly mentioning gender, age (±5 years, based on Flin & Shepherd, 1986), face, name, body, hair, clothes, boots.

The content of answers to the questions were divided into categories depending on the questions asked: gender (male or female), age, height, figure, face, mouth, type of nose, size of nose, shape of nose (straight or hooked nose), ears, eyes, parting of the hair, hair colour, hair length, upper body clothing and colour, lower body clothing and colour, boots, colour of boots. The rules for scoring answers to the questions were one point correct answer in cases of gender, age (±5 years), height (±5 cm), figure, face, mouth, type of nose, size of nose, shape of nose (straight or hooked nose), ears, eyes, parting of the hair, hair colour, hair length, boots, colour of boots, type of upper body clothing, colour of upper body clothing, type of lower body clothing and colour of lower body clothing.

In own-anchor effect analysis, the questionnaires about participants' own characteristics were coded. For each characteristic the participants were categorized as being either (i) similar or (ii) not similar to the target. Similar coding system was used as described above.

4.4. RESULTS

First will be presented information concerning the possible influence of participant gender and ethnicity on their 'free' description performance. The 'standard' was not

used during this description stage. Then will be presented the effects of gender, ethnicity, and of using a standard when answering the questions. Then the effects of the typicality and gender of the standard are examined. Also will be presented information concerning the possible influence of participant similarity on their descriptions of the target. Finally, gender differences in these descriptions will be examined.

4.4.1. Descriptions

In their descriptions the young adults mentioned an average of only 5.03 semantic units (but in answers to the questions 15.69 semantic units). A two-way ANOVA for gender and ethnicity indicated that there were no effects of gender or ethnicity on the free descriptions. The descriptions mostly included the characteristics of clothes, hair, and face.

4.4.2. Answers to the questions

For total performance (i.e. in answering all the questions, combined) a three-way ANOVA for (i) gender, (ii) ethnicity, and (iii) use of a standard produced no main effects. Even when the effects of gender, ethnicity and standard were examined for each of the questions separately, no significant effects were found.

4.4.3. The effect of type of standard

A one-way ANOVA examining for the effect of type of standard on answers to the questions revealed a statistically significant effect $F(2,90) = 6.58, p < .01$, Cohen's $d = 1.39$ (the typical female and the typical male standard produced considerably better effects than did the atypical male standard, see Figure 4.1). Further analysis revealed that while there were no significant effects of using the atypical male standard or the typical female standard, there was a tendency for an enhancement effect using the typical male standard $F(1,31) = 3.42, p < 0.07$, Cohen's $d = 1.23$.

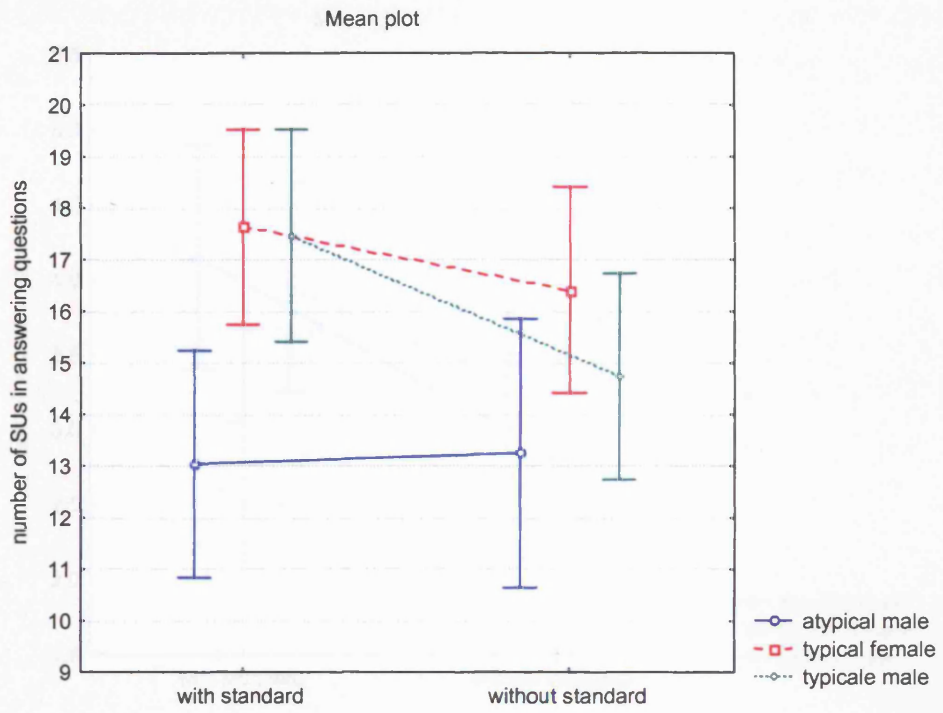


Figure 4.1. The effect of type of standard and no standard in answering questions.

4.4.4. Effect of use of a standard regarding 'easy' and 'difficult' characteristics

According to the previous literature (Kuehn, 1974; Ellis, 1990; Sporer, 1992; Lindsay, Martin, & Weber, 1994; Van Koppen & Lochun, 1997; Pozzulo & Warren, 2003) adults describe better such characteristics as gender, height, age, face, hair style, hair colour and length, upper and lower body clothing and colour; and poorer characteristics such as figure, mouths, nose, nose size, nose hookedness, ears, and eyes. A two-way ANOVA for the type of standard and 'easy' vs 'difficult' characteristics (based on the previous literature) was conducted which revealed a significant interaction $F(1,91) = 5.79$, $p < .02$, Cohen's $d = 1.21$). Further analysis revealed there was a significant effect of using the typical male standard for 'difficult' characteristics ($F(1,31) = 4.95$, $p < .03$, Cohen's $d = 1.78$) (Figure 4.2) but not for 'easy' characteristics (Figure 4.3).

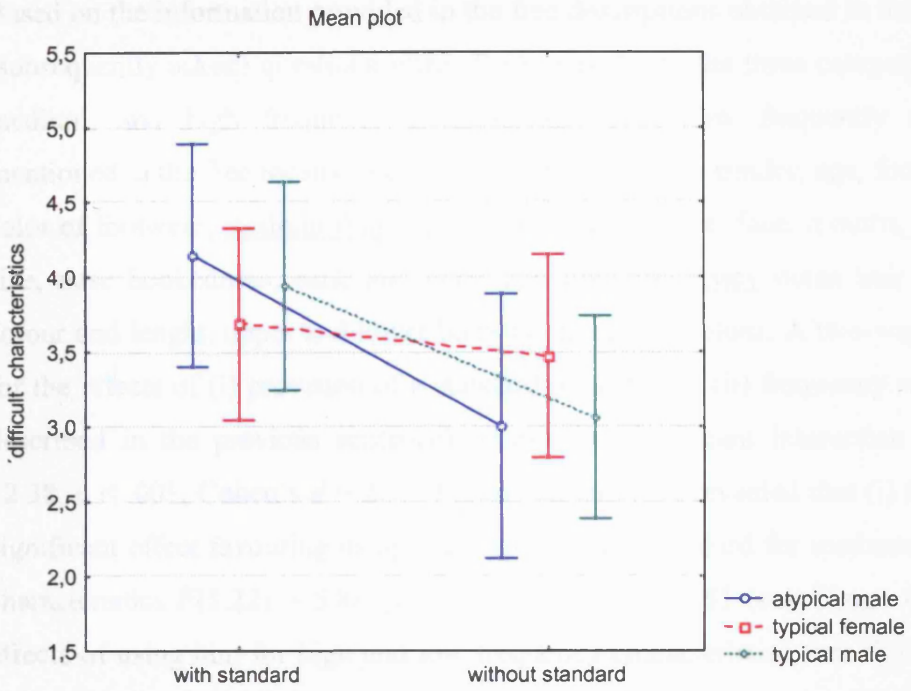


Figure 4.2. The effect of type of standard and no standard for ´difficult` characteristics.

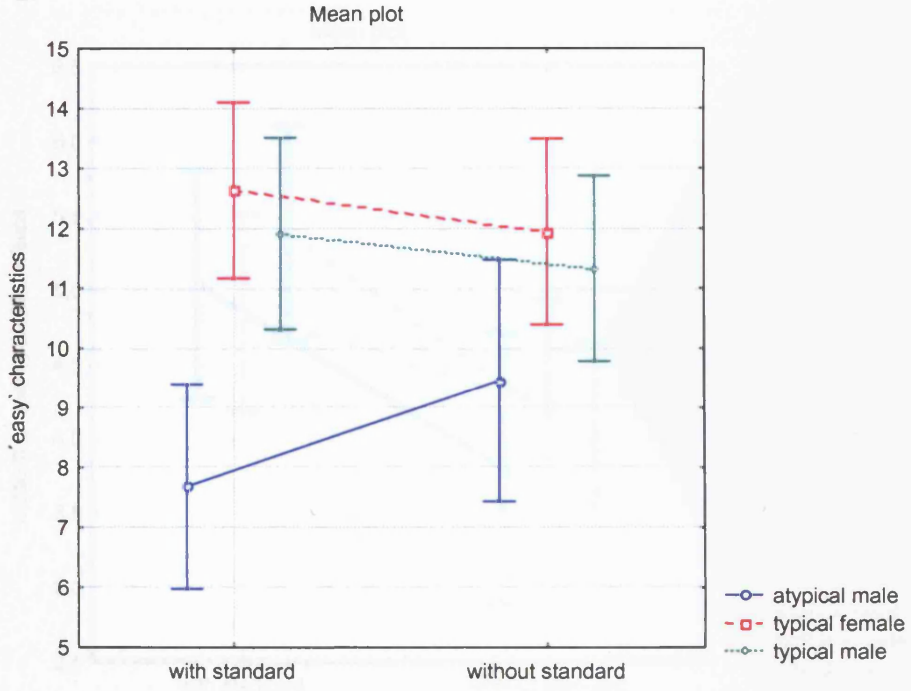


Figure 4.3. The effect of type of standard and no standard for ´easy` characteristics.

Based on the information provided in the free descriptions obtained in this study the (subsequently asked) questions were also grouped into the three categories of low, medium, and high frequency characteristics (i.e. how frequently they were mentioned in the free recalls). Low frequency items were gender, age, footwear, and color of footwear; medium frequency items height, figure, face, mouths, nose, nose size, nose hookedness, ears, and eyes; and high frequency items hair style, hair colour and lenght, upper and lower body clothing, and colour. A two-way ANOVA for the effects of (i) provision of a standard (or not) and (ii) frequency revealed (as described in the previous sentence) revealed a significant interaction $F(1,91) = 12.38$, $p < .001$, Cohen's $d = 2.59$. Follow up analyzes revealed that (i) there was a significant effect favouring using the atypical male standard for medium frequency characteristics $F(1,22) = 5.88$, $p < .02$, Cohen's $d = 2.51$ (see Figure 4.4) but no effects of using him for high and low frequency characteristics (see Figure 4.5 and 4.6), (ii) there was also a significant effect of using the typical male standard for medium frequency characteristics $F(1,31) = 10.79$, $p < .001$, Cohen's $d = 3.88$ but no effects of using the typical female standard.

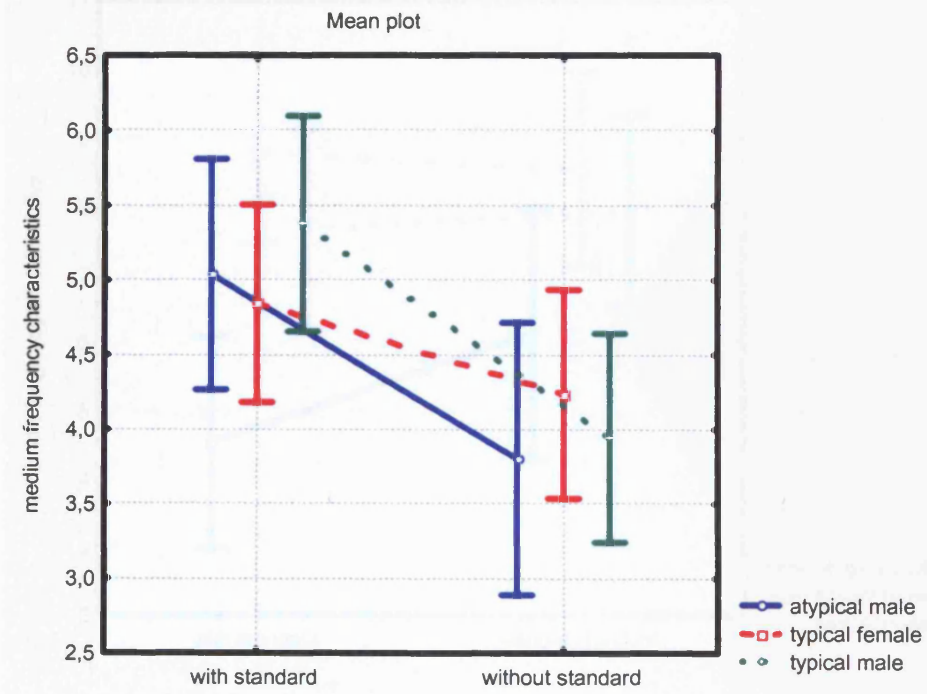


Figure 4.4. The effect of type of standard and no standard in medium frequency characteristics.

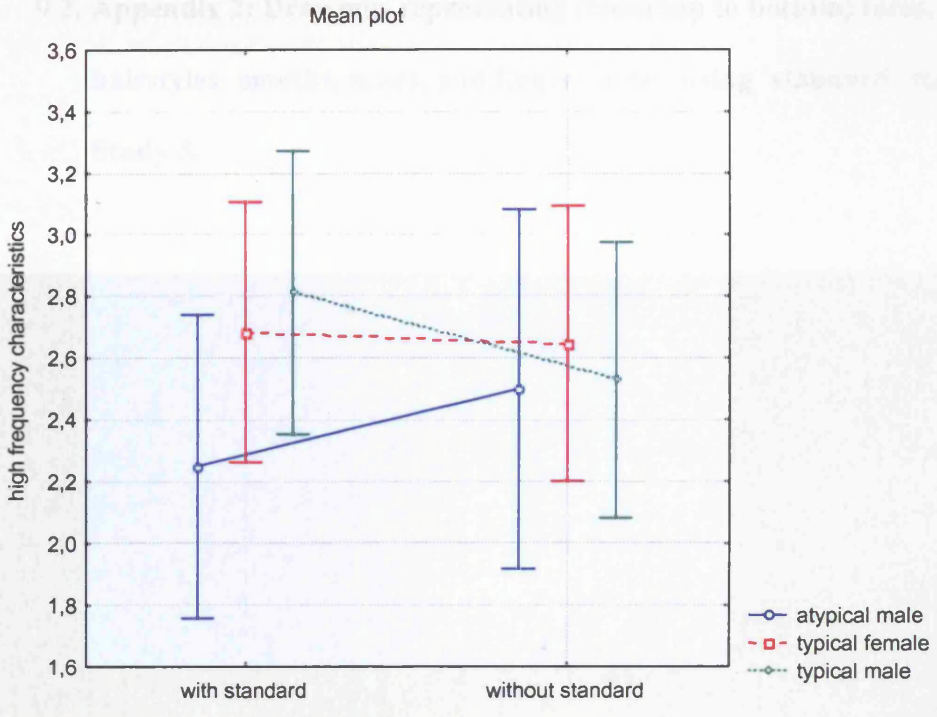


Figure 4.5. The effect of type of standard and no standard in high frequency characteristics.

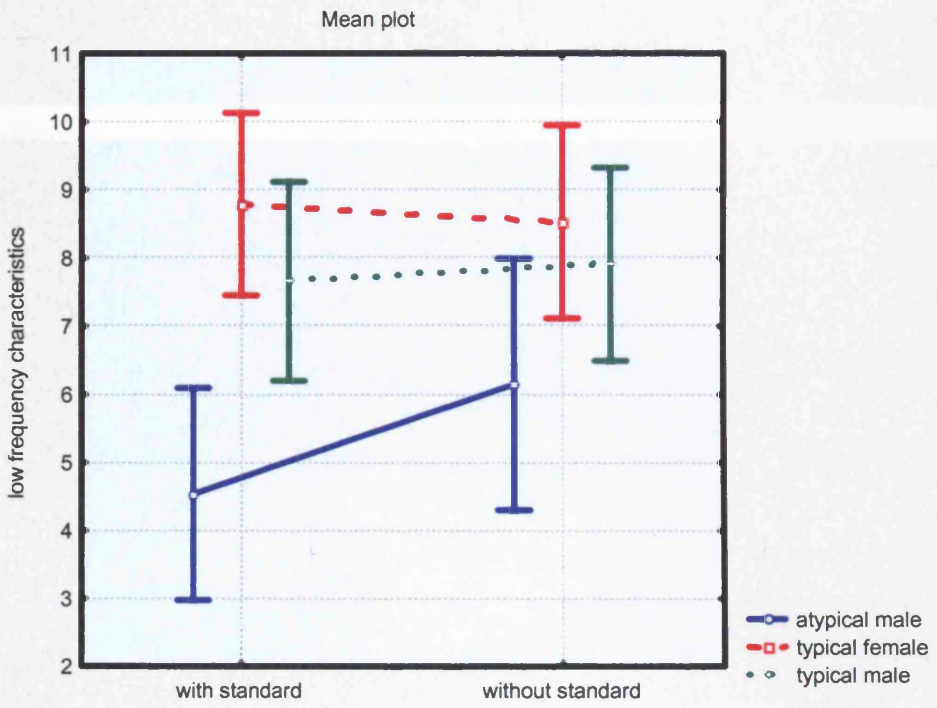
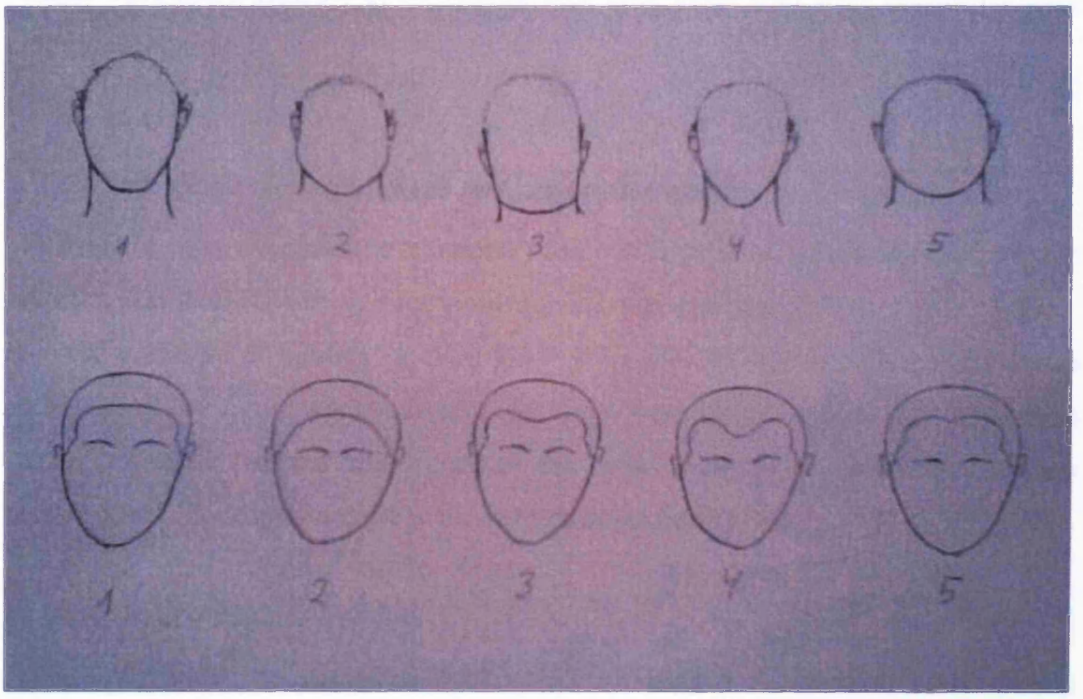
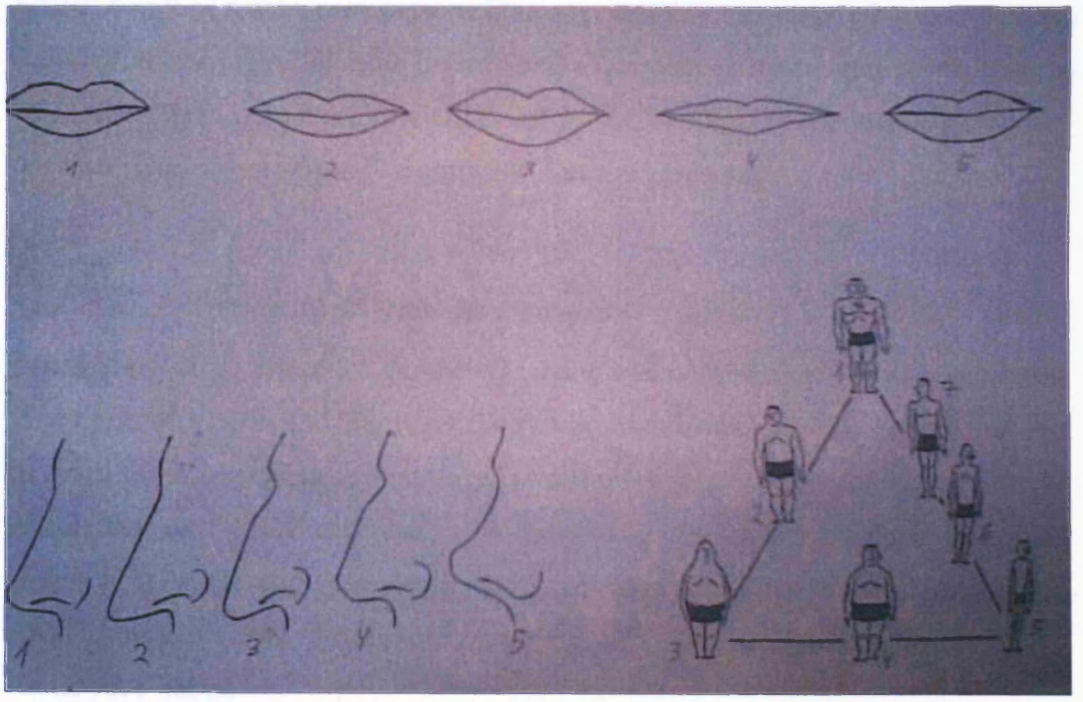


Figure 4.6. The effect of type of standard and no standard in low frequency characteristics.

9.2. Appendix 2: Drawings representing (from top to bottom) faces, hairstyles, mouths, noses, and figures in the using ´standard` mode in Study 3.





4.4.5. Own-anchor effect

Now own-anchor effects on person descriptions are presented. Chi-square tests were used to examine the possible effects of witnesses' several characteristics on similarity and accuracy. A significant effect emerged for the shape of face $\chi^2(1, N = 92) = 4.24$, $p < .05$ indicating that participants whose face shape was similar to the target were more accurate at estimating face (and those who were dissimilar were more inaccurate).

4.4.6. Gender differences on own-anchor effects

Chi-square tests were used to examine the possible role of gender on own-anchor effects. Main effects emerged for males for the shape of face $\chi^2(1, N = 92) = 5.51$, $p < .05$ and type of figure $\chi^2(1, N = 92) = 4.20$, $p < .05$ which indicate that males whose face and figure were similar to the target were more accurate (and those who were dissimilar by face and figure to the target were more inaccurate in their estimations). For females there were no significant main effects.

4.5. DISCUSSION

This study found no straightforward effect of using a standard to assist young adults' recall of a once-seen person. This suggests that the usage of standard to aid (child) witnesses to recall more information about once-seen stranger is not likely to work. Also there were no simple effects of gender or ethnicity. Concerning the own-anchor effect, only a very few significant findings emerged.

4.5.1. Young adults' face descriptions

The finding of no effects of gender on young adults' free descriptions contradicts Powers, et al. (1979) and Yarmey (1993) who found women to be superior to men in recall for characteristics about target's hair and figure. Also, there were no ethnic differences in the present study (for describing the Caucasian persons) which supports Fallshore and Schoolers' (1995) findings that there were no significant relationships between description accuracy and identification performance for 'other-race' faces.

In present study it was found that the young adults' free recall focused more on clothes, hair and face, as Sporer (1992b) and Lindsay et al. (1994) have found. However, these characteristics are the most changeable (compared to the stable characteristics such as facial features, figure, height or weight). Similarly to previous studies, the amount of information produced by young adults in their person descriptions was relatively low.

4.5.2. Using the standard in answers to the questions.

In answering the questions about the target, there were also no significant effects of gender or ethnicity. There was no simple overall effect found for using a standard to help to recall characteristics of once-seen person, which is similar to Kask et al.'s (2007) findings of no effect of a standard on improving children's person descriptions. The findings of this study and Kask et al. suggest that using a standard does not help.

However, there was a tendency for the use of a typical male standard to assist, which was significant for characteristics that the previous literature suggests are not spontaneously mentioned in free descriptions such as figure, mouths, nose, nose size, nose hookedness, ears, and eyes. The possible effect of typicality of the standard is not so clear yet; therefore this effect could be studied further.

4.5.3. Own-anchor effect

Participants who were similar to the male target were more correct in estimating inner face characteristics such as face and nose hookedness. Gender differences regarding own-anchor effect were found for males but not for females. For males, there was an own-anchor effect present for face and figure characteristics, namely, participants who were similar to the target in these characteristics, were more accurate in their estimations than those who were more dissimilar to the target. These findings only partly support the own-anchor effect. However, as the effects are weak, we cannot say that person's own characteristics are important when estimating the other persons' characteristics.

4.6. CONCLUSION

The use of standard did not help young adults to recall more characteristics about the once-seen person (which was also found previously for children). Also no major own-anchor effect was found in this study. The factors which might have been influenced the current findings were that participants were tested in a group and therefore some of the participants might have not initially noted all the target's characteristics. (Additional experiments with individual testing may be beneficial to provide better visibility of the target.) However, in real life many crimes take place when people are not fully focusing on the perpetrator's characteristics. Also, as the period of time participants saw the target was quite short (one minute), an experiment employing a longer period of time could be used.

Further research should address methods to increase the information about 'difficult' characteristics. Also, the typicality of the standard should be studied further.

5. CHAPTER 5: ESTONIAN CHILDREN'S AND YOUNG ADULTS' CROSS-RACE FACE RECOGNITION ABILITY.

5.1. ABSTRACT

Purpose. There appear to have been published no studies of the cross-racial effect with multi-ethnic targets in less 'Westernised' societies, such as in Estonia.

Method. Estonian children's (N=65) and young adults' (N=62) face recognition ability for different race faces is examined. A new target presentation method was used where participants were shown a stimulus face sheet of four different race faces. Then a sequential larger face recognition set was used in which some of the four stimulus faces were present and some absent. Participants had to make a choice for each face.

Results. Own-race faces were more correctly rejected than other-race faces. For other-race faces correct recognition decreased when similar foils were presented before targets in the recognition set but this did not occur for own-race faces.

Conclusions. The findings suggest that when a multi-perpetrator crime takes place, then identifying the real culprits can be problematic and will depend on different aspects of the face recognition task such as own- or other-race targets; target present or absent; and/or the location of foils and target(s) in the lineup.

5.2. INTRODUCTION

In spring of 2006 there was a letter sent all over the University of Leicester campus to warn especially females that two men have tried repeatedly to kidnap females near the university campus area. One of the men was described as being East-European and the other one as South-Asian.

There are numerous of studies on face recognition of different race faces which have largely found evidence for the ´cross-race` effect (or ´own-race` bias), namely people recognizing their own-race faces better than other-race faces (Bothwell, Brigham, & Malpass, 1989; Chance & Goldstein, 1996; Meissner & Brigham, 2001; Pezdek, Blandon-Gitlin, & Moore, 2003; Sporer, 2001a; Wright, Boyd, & Tredoux, 2001; Shapiro & Penrod, 1986). This cross-race recognition effect has been found in field settings (Brigham, Maass, Snyder, & Spaulding, 1982; Platz & Hosch, 1988) as well as in laboratory-based studies (see Meissner & Brigham, 2001, Sporer, 2001a; 2001b for a review). Such studies have suggested that individuals devote more processing to features relevant to own-race faces. However, Anthony, Copper, and Mullen (1992) found that the cross-race effect among Caucasian participants was stronger than among Black participants and Chiroro and Valentine (1995) found no ´cross-race` effect among Black African participants.

The majority of the previously published cross-racial face recognition studies have been conducted in parts of highly ´Westernised` societies, namely the United States, United Kingdom, South Africa, and Germany, each with a history of recent migration or segregated cultures. There appear to have been published no studies of the ´cross-race` effect in less ´westernised` societies. The present study examines whether the ´cross-race` effect is also present (for both children and adults) in a society which has been more closed (for a long time) to ethnic migration (i.e. Estonia). Estonia is a homogeneous society in which the people are predominantly Caucasian. It was until recently a ´closed` society to other-race migration, ´belonging` to Soviet Union for almost 50 years. However, during the past 15 years Estonia has been an independent country. There is no major issue of migration yet, and there still is a rather limited day-to-day direct exposure to other race faces.

5.2.1. Different mechanisms in cross-race face recognition

One notion initially put forward as an explanation of the cross-race effect was that of physiognomic homogeneity (i.e. faces from some races might actually show less physiognomic variability among group members when compared with other races). However, this has not been found to be valid (Goldstein, 1979; Goldstein & Chance, 1979), suggesting that a more complex explanation is necessary.

5.2.2. Social and cognitive mechanisms in the cross-race effect

A number of psychological explanations for this phenomenon have been suggested but there is little consensus regarding the actual social and cognitive mechanisms that may govern the cross-race effect. Early publications on this topic mentioned the possible effects of social attitudes and recent hypotheses have involved the possible influence of interracial contact and the notion of perceptual learning (Meissner & Brigham, 2001).

Early research on the cross-race effect demonstrated a small relationship between attitudes toward other-race persons and recognition performance (Berger, 1969; Galper, 1973) in that those Caucasian participants whose attitudes towards Afro-Americans were negative recognised Afro-American faces poorer than did those who did had neutral attitudes. However, more recent studies have failed to find a relationship between racial attitudes and memory for other-race faces (Slone, Brigham, & Meissner, 2000; Swope, 1994; Meissner & Brigham, 2001). Indeed, in their meta-analysis Brigham and Meissner (2001) found that racial attitudes appeared to have no direct influence on the cross-race effect. A possible mediating role was indicated by a moderately strong relationship between racial attitudes and interracial contact in that participants with more prejudiced attitudes have reported less contact with other-race members (Brigham & Meissner, 2000; Slone et al., 2000).

Thus, the quality or quantity of interracial contact may play a role in the cross-race effect. Meissner and Brigham (2001) proposed that increased contact with other-race individuals may reduce the cross-race effect by (i) reducing the likelihood of stereotypic responses (and thus increasing the likelihood that individuals may look

for more individuating information [Malpass, 1981]), and/or (ii) influencing individuals' motivation to accurately recognize other-race persons though associated social rewards and punishments (Malpass, 1990), and/or (iii) reducing the perceived similarity (or complexity) of unfamiliar other-race faces (Goldstein & Chance, 1971).

Some studies of the relationship between recognition of other-race faces and individuals' self-reported experience with other-race persons have found significant relationships (Lindsay, Jack, & Christian, 1991; Slone et al., 2000; Swope, 1994; Wright, Boyd, & Tredoux, 1999; Platz & Hosch, 1988). Carroo (1986) found evidence of the influence of inter-racial contact in Caucasian and Black samples from UK and Africa (Carroo, 1986). Sporer (1999) found a weak but significant cross-race contact effect for White German participants with Turkish faces, but not for Turkish participants living in Germany tested on German White faces. However, some studies have found no effect of experience (Berger, 1969; Cross, Cross, & Daly, 1971; Burgess, 1997; Ng & Lindsay, 1994).

In a recent study Walker and Hewstone (2006) found that British White participants were better at discriminating White faces than South Asian faces, whereas British South Asian participants were equally accurate in discriminating both type of faces. They concluded that a lack of other-race individuating experience can predict an own-race effect not simply at the stage of recognition but also at the stage of perceptual encoding. Chance and Goldstein (1996) proposed that a 'cohort effect' may exist, namely that opportunities for interracial contact have increased in some countries (and that the resultant less strong racial attitudes may now have a weaker influence).

One explanatory approach in this field is that of perceptual learning which has been defined as "an increase in the ability to extract information from the environment, as a result of practice and experience with stimulation coming from it" (Gibson, 1969, p. 3). Thus, people may discriminate own-race faces more accurately because they use more appropriate aspects of the face. However, cues used for own-race faces may not be appropriate for other-race faces. Diamond and Carey (1986) proposed a configural–featural hypothesis regarding facial recognition, according to which

experienced participants rely on configural aspects whereas novel participants rely on featural aspects of the face (see also Farah, et al., 1998). Relatedly, Rhodes, et al. (1989) proposed that greater experience with own-race faces would lead to a larger inversion effect (which is a situation when the recognition of face pictures are disproportionately affected by a 180 deg rotation in the image plane from the normal, upright viewing condition) due to an increased reliance on configural information, whereas encoding of other-race faces should not be so influenced by inversion due to reliance on featural aspects. However, several studies have noted no interaction of inversion with the cross-race effect (Buckhout & Regan, 1988) and others a larger inversion effect for other-race faces (Valentine & Bruce, 1986).

Valentine and his colleagues (Valentine, 1991; Valentine & Bruce, 1986; Valentine & Endo, 1992) have proposed an exemplar-based model of facial memory reflecting "the acquisition of knowledge of how faces vary" (Valentine, 1988, p. 485). This model may account for the effects of inversion, race, and distinctiveness. His multi-dimensional space (MDS) framework states that with increasing experience, other-race faces may be better represented once the relevant aspects of other-race faces are learned. Chiroro and Valentine (1995) examined the effects of race, typicality, and level of perceptual experience within the cross-race paradigm. They predicted and found that participants who had considerable previous experience with other-race faces (high-contact) demonstrated better distinctiveness effects (i.e. recognizing both own- and other-race faces equally well) and that low-contact participants (who do not have much experience with other-race faces) showed worse distinctiveness effects (i.e. recognizing correctly own-race faces but not other-race faces).

Levin (1996) noted a paradoxical effect in that individuals are slower at classifying the race of own-race faces compared to other-race faces, and termed this the 'other-race classification advantage' (ORCA). Levin suggested that other-race faces were more quickly classified due to an automated process in which race-specific coding is performed without regard for other individuating information.

Sporer (2001a) demonstrated an in-group/out-group model of face processing (IOM, see Figure 5.1) according to which for in-group faces (the default) automatic

processing commences with configural coding (characteristic of expert processing of a normal upright face), whereas with out-group faces the out-group characterization cue/s/ automatically start to trigger categorization before other (the default) face-processing strategies. In other words, in-group faces are processed more holistically and automatically whereas out-group faces are first categorized on the basis of some feature specific to that particular group and then less efficient encoding strategies are applied such as relying on feature-oriented instead of more efficient holistic encoding (Sporer, Trinkl, & Guberova, 2007). According to Sporer, although the initial categorization may be useful and efficient for classifying face belonging to a particular out-group, it is considered less appropriate when the task involves individuating a particular face from other members of that group. Also, people encode insufficient details for discriminating between members of this out-group.

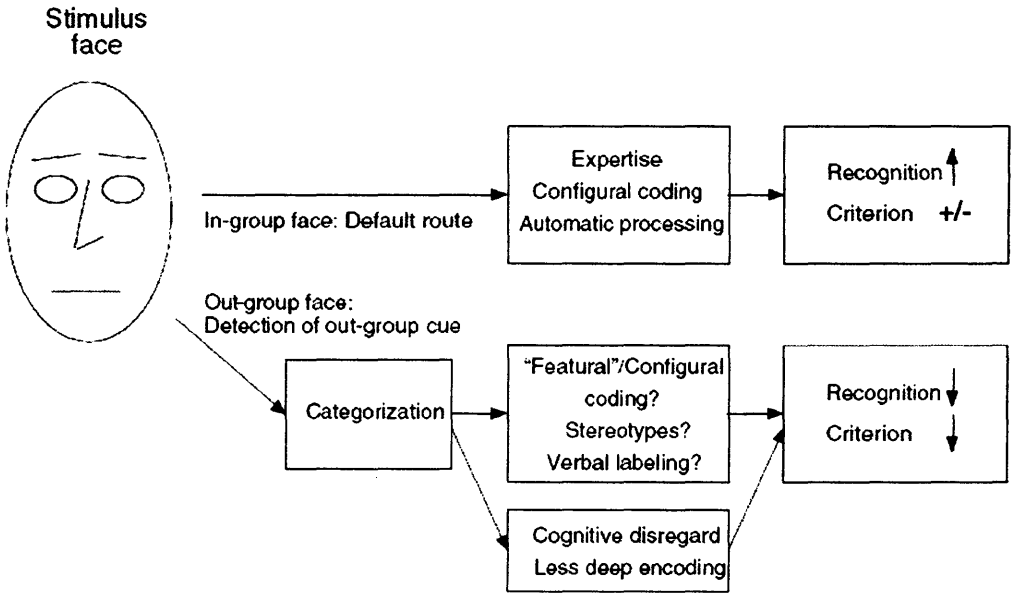


Figure 5.1. In-group/out-group model of face processing (Sporer, 2001a). Image from Sporer (2007).

Doyle (2001) proposed than White American eyewitnesses might be more willing to guess when attempting to identify Black people from lineups than when attempting to identify White people from lineups. Meissner and Brigham (2001) found support to this notion, namely, participant-witnesses appear to use a more liberal criterion with other-race faces than with own-race faces. It means that this 'accidental' or

'biased' identification of other-race faces (regardless of being accurate or not) can be a serious problem to address in people's person identification. However, Doyle did not state that this effect would occur in the 'opposite' direction, i.e. Black witnesses are more willing to guess when identifying White American faces (see also Wells & Olson, 2001).

5.2.3. Children's face recognition ability and cross-race effect

The ability to recognize faces increases with age (Ellis, 1990; Hay & Cox, 2000), more rapidly improving between age of six and ten than later (Davies, 1996). Carey (1992) noted that there is a little difference between ten-year-olds and adults. With age children's ability to remember distinctive features of faces improves, which helps them subsequently to discriminate previously seen faces from other faces (Ellis, 1990).

According to Goodman et al. (2007), the cross-race effect may be more plastic for children than for adults, especially in pre-school years. They noted that it is possible that the cross-race effect is evident for young children in some but not in other circumstances. The development of the cross-race effect could depend on factors such as depth of processing, encoding time, and configural/featural processing. Longer, deeper (e.g., semantic), and configural (e.g., relations among facial features) encoding, as opposed to shorter, shallower (e.g., physical attributes), and featural (e.g., face shape, eye size) encoding is associated with better face recognition. They concluded that regarding featural/configural processing, featural encoding could explain the cross-racial effect if people attend to less informative features in cross-race faces than in own-race faces (see also Brigham et al., 2007).

There are not many experiments which have examined whether or not the other-race effect applies for children and the findings seem to be somewhat controversial where more recent studies seeming to find larger own-race effect than did older studies. However, this could be influenced by what ethnicity faces were used in certain experiments.

Goldstein and Chance (1980) tested Caucasian children (from 1st to 6th grade) and adults for the recognition of either Caucasian or Japanese faces, and found that although children's performance improved with age, there were no significant differences in recognition of Caucasian or Japanese faces for children; however, the differences were present in adults who recognized Caucasian faces better. Chance et al. (1982) found larger differences in performance between Caucasian and Japanese faces for older participants than for younger ones (children aged 7 to 8, 11 to 12, and 13 to 14 years). Also, Cross et al. (1971) reported that Caucasian participants (7, 12 and 17 year-olds) were poorer with Black faces than with Caucasian faces, whereas Black participants were equally good with both type of faces.

More recently, Pezdek, Blandon-Gitlin, and Moore (2003) found that, similarly to adults, children recognize own-race faces more accurately than cross-race faces. In a study by Corenblum and Meissner (2006) two experiments demonstrated that children (aged 9 to 12 years) were better at recognizing own-race faces (Euro-American vs African American) and also more faces from their own ingroup (i.e. children recognized better other children's faces). Goodman et al. (2007) studied the cross-race effect in the United States, Norway, and South-Africa for recognition of Asian, African, and Caucasian faces and found that regardless of national background 8 to 14-years-old children and adults recognised own-race faces more accurately than other-race faces whereas five to seven-year-olds recognised all face types equally well. Another cross-race effect experiment with 10 to 15-years-old Turkish and Austrian children demonstrated that Turkish children were more correct than Austrian children with Turkish faces, however, there were no differences in recognising German faces (Sporer, Trinkl, & Guberova, 2007).

These findings support both Valentine's MDS theory and also Sporer's IOM model of face processing where repeated and continuous exposure to face of ingroup (and to a lesser extent to outgroup) members results in knowing and using those dimensions appropriate for processing and recalling faces of people who perceivers know well, a process that occurs early in life and remains relatively stable across a wide age range. Slone et al. (2000) claim that own-group biases in face recognition may be due more to saying 'seen before' to new faces of outgroup members (i.e. false alarm responses) than to accurate recognition of faces of ingroup members.

An effect of interracial contact has been also found in children. Adolescents and children living in integrated neighbourhoods recognized better novel other-race faces than did those lacking of experience with other-race faces (Cross, Cross, & Daly, 1971).

5.2.4. Face description ability

People have considerable experience of recognizing faces, especially familiar faces, but are less experienced at describing faces. Very few studies have attempted to determine whether the way people describe faces differs regarding (i) their own and (ii) other race faces (Sporer, 2001b). Those who have examined descriptions of own- and other-race faces have suggested that individuals attend to features relevant to own-race faces (Shepherd & Deregowski, 1981). However, Fallshore and Schooler (1995) found no cross-race effect in witnesses' verbal descriptions. Thus, while face recognition usually reveals an own-race advantage (at least for Caucasians), face description has generally failed to show such a race effect (Meissner & Brigham, 2001; Sporer, 2001a, 2001b). It is known that there are no relationships between description accuracy and identification performance for both own-race faces (Cutler, Penrod, & Martens, 1987) and for other-race faces (Fallshore & Schooler, 1995).

Ellis, Deregowski, and Shepherd (1975) compared the actual face descriptions by Black African and White British participants. White British participants more frequently mentioned hair colour and texture as well as the colour of the iris of the eye. By contrast, Black Africans provided more descriptions of the ears, chin, eyebrows, size and white of eyes, style of hair, and face shape. Sporer (2001b) concluded that Black faces are described by both Black and White participants by referring to a larger number of facial characteristics. However, Black participants used a larger number of facial characteristics when describing both Black and White faces. In general, he suggested that Black faces are not more difficult to describe than White faces, but this may depend upon the country in which such studies are conducted.

5.2.5. Present study

Estonian children's and young adults' face recognition ability regarding different race faces is examined in this study. It is hypothesized that Estonians' face recognition ability could be similar to those in 'Western' countries (i.e. recognition of own-race faces is better than other-race faces). Estonia is a homogenious society and has been closed for exposure of people in other races for a long period of time. Age differences in own- and other-race face recognition are also examined.

In this study a new type of target presentation is employed in which a small number of different race targets are presented before the multi-race recognition set. This was done because in real life several targets may be seen at the same time but they could be of different races (e.g. when a witness sees four men of different ethnicities committing a crime). Concerning this (sequential) several race face recognition experiment, it is examined whether participants have different criteria when recognizing own- and other-races faces (i.e. is there more liberal criterion involved in other-race face recognition as Doyle (2001) and Meissner and Brigham (2001) proposed). As this notion has largely been overlooked in previous research, this study examines it more closely.

The position of targets and foils in the recognition set is examined. It is hypothesized that targets and foils are chosen more when presented in the beginning than in the end of the recognition set. Finally, young adults' face description ability is also examined.

5.3. METHOD

5.3.1. Participants

The participants were young adults (N=62, with the average age of 17.5 years, range 16 to 19 years) and children (N=65, with the average age of 9 years and 2 months, range 8 to 10 years) who were recruited by notices distributed at schools in Estonia. All participants spoke Estonian as their primary language and were Caucasian.

5.3.2. Design

A 2 (age of participants; child vs adult) × 2 (different race faces; own-race (i.e. Caucasian) vs other-race (i.e. mixture of Afro-American, Latino, Chinese, and Turkish faces)) × 2 (type of presentation; target-present vs target-absent) between-participants design was used.

5.3.3. Procedure

The male experimenter entered a classroom where participants were together with their teacher and introduced himself. Children and young adults were tested in small groups (both age groups had four class groups). First of all, participants were shown a set of four stimulus faces. On each stimulus face sheet there were four faces of different ethnicities chosen from the five ethnicities: Caucasian (N=4), Afro-American (N=4), Turkish (N=1), Latino (N=4), or Chinese (N=2) (see Appendix 3). An own-race face (i.e. Caucasian) was always shown within those four faces. This stimulus face spread sheet was shown to participants for about 30 seconds (see Appendix 4).

Then the participants were handed a short personal questionnaire to complete in order to create a delay between face presentation and recognition. After this the participants were shown a face recognition set. Both of the face recognition sets employed had two practice and thirty test faces which were shown sequentially (15 second each) using an overhead projector (see Appendix 5). The participants had to decide for each face whether it was or was not one of the four faces they had seen earlier. When the participants were not sure about a test face they could indicate 'I don't know'. Two of the four stimulus faces were present among these 30 faces (this option is referred as a target-present dataset), and the other two stimulus faces were not (target-absent dataset). For each target face two faces corresponded in the recognition set (i.e. in the target-present condition the same photograph target and a photograph of one foil were presented and in the target-absent condition the target replacement and one foil were presented).

Then (at the end) the young adults were also asked to describe one of the target faces which was visibly present during this description task. (The children did not

participate in this task.) Person descriptions were collected after the recognition task to avoid the possible presence of verbal overshadowing.

5.3.4. Test faces

The non-target faces in the test set of faces were selected on the basis of 'ratings of similarity of foils to the perpetrator' procedure used by Pigott and Brigham (1985). A pool of 89 (possible test) photographs judged by the current author to be similar to the targets were rated by 18 mock witnesses using a 1 to 5 scale (1 being 'dissimilar' through to 5 being 'extremely similar' for each ethnicity). The photograph with the highest similarity rating to each target was chosen as the target replacement (for the TA recognition set) and two photos with the next highest similarity ratings were chosen as foils for the test set (the similarity rating had a mean of 3.94). Distinctiveness ratings were not obtained for the faces. The test set comprised 30 black-and-white headshots of the face. The targets, target replacements, and foils were randomly placed, across participants in all positions in the test array.

5.3.5. Coding

The contents of young adults' face descriptions were divided into the following categories: gender (male or female), age, face, body, hair, clothes, nose, ears, mouth, eyes, eyebrows, moustache, beard, skin, teeth, race, neck, and accessories. The contents were coded (i.e. scored) for the total number of semantic units (SUs) (using the same method as Poole & White, 1991). The coding was performed by two different coders who were unaware of the purpose of this study. (Inter-coder agreement was 88%.)

5.4. RESULTS

The proportions of correct and false identifications in the target-present dataset and the proportions of correct rejections in the target-absent dataset were computed. First will be presented information concerning age differences regarding own- and other-race face recognition. Then will be examined whether the most similar foil in

the target-present dataset has an effect on performance. Finally, young adults' face descriptions are examined.

The effects for different race face recognition are examined by conducting logistic regression analysis. Other-race faces (Afro-American, Turkish, Latino, and Chinese) are analyzed together to be able to conduct logistic regression (otherwise the cells would be too small to conduct this statistical analysis). Also, chi-square analysis with Fisher's Exact was performed where some cells had an expected count of below five (i.e., analysis concerning presenting target or the most similar foil first or second).

A main aim of this study was to examine the effect of several different ethnicity faces, compared to typical cross-race effect experiments where only own-race and one other-race faces are used. However, as some race faces were not so well-represented in this study, which could have a negative effect on the results and interpretations. Study 5 takes these shortcomings into consideration where a fuller and better-balanced design is used.

5.4.1. Effects of different races.

Table 5.1 presents the percentages of participants (adults and children combined) choosing the target correctly (correct identification) or choosing a foil (false identification) for the target-present dataset; and the percentage of correct rejections for the target-absent dataset.

Table 5.1. *Percentages and Frequencies of Correct Identifications, False Identifications, and Correct Rejections.*

	Own-race faces	*N*	Other-race faces	*N*
Correct identifications	73%	52	83%	212
False identifications	30%	50	36%	210
Correct rejections	74%	112	54%	226

Logistic regression revealed that own-race faces were more often correctly rejected than other-race faces B = .92 (.34), B(Exp) = 2.51, $p < .001$ ($R^2 = .11$ (Cox & Snell),

.16 (Nagelkerke), model $\chi^2(1) = 17.98$, $p < .001$). For correct and false identifications there were no statistically significant race effects.

5.4.2. Age differences in the recognition of different race faces.

The differences in adults and children between own- and other-race face recognition for correct and false identifications, and correct rejections are now compared. Adults more often correctly rejected own-race than other-race faces B = 1.14 (.36), B(Exp) = .32, $p < .01$ ($R^2 = .09$ (Cox & Snell), .12 (Nagelkerke), model $\chi^2(1) = 7.95$, $p < .01$). Similarly, children also more often correctly rejected own-race than other-race faces B = 1.23 (.44), B(Exp) = .28, $p < .01$ ($R^2 = .07$ (Cox & Snell), .10 (Nagelkerke), model $\chi^2(1) = 8.85$, $p < .01$).

There were no significant differences between adults and children concerning correct and false identifications, and correct rejections.

Table 5.2. *The Percentages and Frequencies of Correct Identifications, False Identifications, and Correct Rejections in Children and Adults.*

Age	Own-race faces	*N*	Other-race faces	*N*
Adults				
Correct identifications	72%	25	83%	102
False identifications	20%	25	38%	100
Correct rejections	73%	56	51%	111
Children				
Correct identifications	71%	28	83%	110
False identifications	40%	25	35%	110
Correct rejections	75%	55	56%	115

5.4.3. Choosing target or most similar foil in the test set.

In the test set, if a target (of a particular ethnicity) was present this was sometimes preceded by the most similar foil. The data were analyzed (see Table 5.3) to examine the effect of this. For own-race faces, when the target was presented first 86% of participants chose it correctly and 29% chose incorrectly the most similar foil which was presented sometime after the target (note that some participants chose first the target and then also the most similar foil). When the most similar foil was presented first, 30% of participants chose it compared to 66% of participants who chose correctly the target which was presented sometime after this foil.

Table 5.3. *Percentages and Frequencies of Choosing Target or Foil First.*

Faces	Target 1st	Foil 2nd	Foil 1st	Target 2nd
Own-race	86%	29%	30%	66%
N	14	14	33	41
Other-race	80%	25%	51%	87%
N	103	109	71	71

There were no differences between targets and foils concerning own-race faces. However, for other race faces chi-square analysis with Fisher's Exact revealed that when foil was presented first, more participants chose it compared to foil being presented second $\chi^2(1, N = 180) = 12.71, p < .001$ (1-tailed).

5.4.4. Age differences in choosing target or most similar foil in test set.

For adults, there were no significant differences in choosing target or most similar foil first or second in the recognition set (see Table 5.4). However, for children own-race target was chosen better when presented first than second $\chi^2(1, N = 33) = 4.55$, $p < .035$ (1-tailed). For other-race faces, when the most similar foil was presented first then it was correctly identified better compared to foil presented second $\chi^2(1, N = 97) = 12.01, p < .001$ (1-tailed).

Table 5.4. *Percentages and Frequencies of Adults and Children Choosing Target or Foil First.*

Age	Target 1st	Foil 2nd	Foil 1st	Target 2nd
Adults				
Own-race	67%	0%	25%	75%
N	6	6	16	16
Other-race	82%	31%	47%	93%
N	48	49	34	28
Children				
Own-race	100%	50%	35%	60%
N	8	8	17	25
Other-race	76%	20%	54%	84%
N	55	60	37	43

5.4.5. Young adults' face descriptions

Finally, descriptions of the different race faces are examined. ANOVA revealed that the effect of race on the size of the descriptions of the faces was not statistically significant, the mean number of SUs for own-race faces was $M = 10.88$ ($SD = 3.56$) and for other-race faces $M = 10.30$ ($SD = 2.60$).

There were differences between the different races only for two descriptor items: (i) for *faces* (i.e. more different features were used in describing person's face) $F(1,61) = 6.41$, $p <. 02$, Cohen's $d = .67$, $M = 2.16$ ($SD = 1.42$) SUs for own-race faces and $M = 1.31$ ($SD = 1.09$) SUs for other-race faces; and (ii) for *noses*, $F(1,61) = 5.84$, $p < .02$, Cohen's $d = .66$, $M = 1.44$ ($SD = 0.92$) SUs for own-race faces and $M = 0.88$ ($SD = 0.78$) SUs for other race faces.

5.5. DISCUSSION

5.5.1. Differences between face recognition of different races.

In this study the 'cross-race' effect was examined in Estonia which is a country what has been more closed to ethnic migration. Also, a new type of target presentation was studied in which a small number of different race targets are

presented before the multi-race recognition set. Own-race faces were recognised as well as were other-race faces. However, for own-race faces participants made more correct rejections than for other-race faces. The position of the target in the lineup appeared also to be significant concerning other-race faces and especially for children, indicating that the placement of the target and foils in a multi-perpetrator lineup (especially with perpetrators being from different ethnicities) is an important issue.

Own-race faces were not more often currently recognized than other-race faces but were more correctly rejected than other race faces. Such findings do not support the notion that people are better at recognizing own-race faces in terms of correct identifications but do so to a certain extent in terms of correct rejections which support partly Levin's 'other-race classification advantage' (1996). Thus, the hypothesis that recognition of own-race faces is better than other-race faces was not supported. Also, there is not a more liberal criterion involved in other-race face recognition which contradicts Doyle (2001) and Meissner and Brigham (2001).

It could be that the relatively high correct recognition rates found in the present study were achieved due to the method used to conduct the study (e.g. stimuli and recognition set was shown on photospreads and not via video or live presentations). Rachlew (2007) in his cross-race effect study found that participants scored significantly higher when the recognition set was presented on photospreads (and not by live or videotaped version). It should be also noted that cross-race effect has been found in some studies when there was a change of view between presentation and test set but not when faces were shown in identical views such as in the present study (and in Study 5) which contradicts previous research (Argstatter et al. 2002 and Gehrke & Sporer, 2006, see in Sporer, et al., 2007). Also, as the stimulus faces in this experiment were identical to those in recognition set then the results could have been influenced by this fact (i.e. being more a picture than face recognition task).

Another factor could influence the results, namely that distinctiveness ratings were not obtained on the faces. It is important to rate distinctiveness of the faces because it is found that people tend to remember more those faces which are more

distinctive (Valentine & Bruce, 1986). Distinctiveness ratings should be conducted by the persons from the same ethnicity, i.e. Caucasians should rate the distinctiveness of Caucasian faces because if Chinese would rate Caucasian faces then for them different characteristics could apply which makes face distinctive to them but not for Caucasians. However, as in this study the other-race faces were not analyzed separately but combined together then it should make the 'distinctiveness' effect smaller (i.e. distinctiveness has a smaller effect when other-race faces are combined compared to single other-race effects). Nevertheless, in some cases other-race faces can be more distinctive due to the factor of novelty and therefore easier to be remembered (which supports Valentine's MDS theory, Valentine, 1991).

Wells and Luus (1990) suggested that correct identifications are largely the result cognitive processes whereas correct rejections also involve social (as well as cognitive) factors. Correctly rejecting faces is more difficult (especially for children) than correctly identifying a face (Pozzulo & Lindsay, 1998). The findings of the present study are similar to those of Pozzulo and Lindsay's, and indicate that the mode of processing could be different between correctly (i) recognizing and (ii) rejecting own- and other-race faces.

The present study found an own-race effect for children which confirms (i) 'encoding switch hypothesis' according to which children encode faces holistically whereas adults do so featurally (Carey & Diamond, 1977; Diamond & Carey, 1977; Diamond & Carey, 1986; Chung & Thomson, 1995) and (ii) other recent findings of the presence of cross-race effect in children (Pezdek, Blandon-Gitlin, & Moore, 2003; Corenblum & Meissner, 2006; Sangrigoli, Pallier, Argenti, Ventureyre, & de Schonen, 2005; Sporer, Trinkl, & Guberova, 2007). Because there were not many differences between nine-year-old children and adults in the present study, these results support the notion that children used more configural than featural face processing strategies. Therefore, it is suggested that development in face encoding ability indeed occurs between infancy and 10 to 12 years (Sangrigoli & de Schonen, 2004) although some younger children could still be more likely to process faces based on featural information despite of being capable of configural processing (e.g., Pellicano & Rhodes, 2003).

It is interesting that some early studies in this field did not find any cross-race effect (Goldstein & Chance, 1980; Chance et al., 1982), which could be due to those studies compared only Caucasian and Japanese faces and that such faces are more similar to each other than Caucasian and Afro-American faces (see Cross et al., 1971). Therefore, the advantage of the current study is that faces from several ethnicities were used.

The position of targets and foils in the recognition set was also examined in this study. When the other-race foil was presented first then participants were more prone to choose it, especially children. Children also chose more when own-race was presented first. Therefore, the hypothesis that targets and foils are chosen more when presented in the beginning than in the end of the recognition set was supported. The findings also partly support the notion that children are more prone to choose someone from a lineup than are adults. Finally, when a most similar foil from an other race was presented before the target, many participants first chose the foil and then also chose the target. However, for own-race faces the target was better recognized than the most similar foil regardless of its position in the test set.

Clark and Davey (2005) showed in their research that the most similar-looking foil was chosen at a higher rate if he appeared in the order late rather than earlier. They invoked a within lineup criterion shift explanation, where when the next best alternative is presented early in the lineup then participants withheld making a choice in order to find out whether a better option would be presented later. In contrast, if low similarity foils appeared first and the next best alternative was in the end of the lineup, participants lowered their decision criterion and chose him. Taken together, these finding indicate that the position of the suspect in the recognition set can affect accuracy.

In this study, to increase ecological validity regarding the initial seeing of several targets of different races, the participants had the opportunity to make multiple choices in a sequential presentation which ended in a way described above (participants tended to choose both foil and target when the target was presented second, especially for other-race faces). These findings are important regarding sequential presentation. When the target is absent in a sequential presentation or is

near to the end of the set, the probability, especially for children, that participants choose the first person who is similar to the culprit (if the lineup is well constructed) is relatively high (especially when the target is from another race). This notion should be studied more closely in further research.

5.5.2. Young adults' face descriptions

Few studies have examined descriptions of own- and other-race faces. In the present study no effect of race was found regarding the total number of descriptors provided. For the various descriptor items, there were only differences between own- and other-race faces for faces and noses. This is in line with previous findings (Fallshore & Schooler, 1995; Meissner & Brigham, 2001; Sporer, 2001a, 2001b).

5.6. CONCLUSION

The present study highlights some of the difficulties concerning the recognition of other race faces. There appears to be shift in processing between (i) correct identification and (ii) rejection. The findings of this study suggest that the correct identification of different race faces differs from correct rejection.

In this study differences between children and adults did not occur for correct identifications or for correct rejections for both own- and other-race faces. In 1986 Shapiro and Penrod in their meta-analysis said that "researchers should devote as much attention to false alarms as they do to hit rates" (p. 152). Meissner and Brigham (2001) supported this notion claiming that (over the past three decades) own-race preference bias appears to have become more prominent in false alarm responses than in correct identifications. The results of the current study also demonstrate that the differences in children and adults do not occur so much for correct identifications, but for correct rejections regarding other race faces.

When a multi-perpetrator crime takes place, then identifying the real culprits can be problematic and will depend on different aspects of face recognition such as own- or other-race targets; target present or absent; and/or the location of foils and target(s) in the lineup. If the target is near to the beginning of the test set then both children and adults correctly identify it. However, when the target is near to the end of the

set, then the number of correct identifications decreases and the number of false alarms increases, especially in children.

The experimental design was suitable for addressing research questions despite of its limitations. A major limitation of this study is that there were too few participants in each ethnicity group that is because this study examined only the differences between own- and other-race faces. Future research should involve more participants in each ethnicity group. Also, as participants in this study were forced to choose (i.e. make the decision about each face although they could use the ´don't know` option), this could have influenced the high rate of false alarms in this study. Therefore, additional experiments should be conducted to test this notion further and to construct a method which on one hand would have good correct identification rates and on other hand would decrease false alarms and incorrect rejections.

Future research should address more precisely the advantages and disadvantages of the new innovative and important procedure employed in this study (i.e. several targets of different races and a multi-race test set). Also, it could be beneficial to replicate this kind of study using a population which could have more expertise in other-race face recognition (e.g. border guards).

6. CHAPTER 6: A COMPARISON OF THREE SEQUENTIAL PRESENTATION METHODS REGARDING OTHER-RACE FACE RECOGNITION

6.1. ABSTRACT

Purpose. In this study three different sequential presentation methods were compared regarding own- and other-race face recognition.

Methods. Estonian young adults' (N=225) were shown a stimulus face sheet of four different race faces. Then a larger sequential face recognition set was used in which some of the four stimulus faces were present and some absent. There were three different methods used concerning how these sequential recognition sets were presented: (i) make decision for each of the face in sequential recognition set (´decide all`); (ii) make decision for only those faces which have been seen before in sequential recognition set (´decide some`); or (iii) seeing consecutive sequential lineups (´sequential lineups`).

Results. Own-race faces were more often correctly rejected than other-race faces. Males identified more correctly own-races faces than did females whereas in correct recognition of other-race faces females outperformed males. There were no effects of presentation method on the recognition of own-race faces. However, there were such effects for other-race faces, namely other-race faces were more falsely identified and less correctly rejected with ´decide all` and ´decide some` presentation methods than with the ´sequential lineups`.

Conclusions. The findings confirm some of the findings of Study 4 where own-race bias was found to positively influence correct rejections and negatively correct identifications. The results reveal that identifying multi-race perpetrators using different presentation methods is a challenge and should be studied further to achieve the best result in different-race face recognition situations.

6.2. INTRODUCTION

The concerns of eyewitness research have always included the improvement of the investigative tools available to law enforcement (Malpass, 2006). Part of this focus is on a reduction in identification errors through improved procedures and personnel training. However, to date there has been little such research on multi-perpetrator crimes even though these are on the rise. For example, homicides involving multiple culprits in the United States have increased since the late 1980s / early 1990s (Bureau of Justice Statistics, 2002), and similarly in Canada (Statistics Canada, 2001). Research so far on this topic has found that when the number of culprits increases, accurate eyewitness identification rates decrease (Shapiro & Penrod, 1986). Considering the growth of multi-culprit crimes (which could involve perpetrators from different ethnicities), more empirical research is needed in this area to help devise effective identification methods.

6.2.1. Sequential lineups

One of the most used presentation methods in eyewitness research is the sequential lineup where the faces are shown to the witness one at a time and the witness has to make a decision for each face whether or not it is the target. This is supposed to diminish the effect of using a relative judgment strategy and rather urge witnesses to use an absolute judgment strategy (i.e. whether <u>this</u> was the target I saw before or not). Lindsay and Wells (1985) found that adults made significantly fewer false alarms in sequential lineups (18.3%) compared to simultaneous lineups (35.0%) when the target was absent (the correct identification rate was not lower for sequential than for simultaneous lineup when target was present - 50% vs 58%).

Steblay, et al.'s (2001) meta-analysis (see a fuller overview in Chapter 5) found that sequential lineups are more efficient when the target is absent, where this presentation method was found not only to increase the number of correct rejections but also decrease the number of correct identifications and false alarms (compared to simultaneous lineups). However, studies with elderly witnesses and children have revealed 'problems' for them using sequential lineups (Wilcock, Bull, & Vrij, 2005).

6.2.2. Multiple choice presentation methods

Given the instructions for sequential lineups, witnesses may expect that they have only one choice and therefore they may select a similar-looking foil before the culprit is presented and thus fail to identify the culprit. If this is so, then allowing witnesses to choose more than once from sequential lineups could help solve the problem of such lost correct identifications. Witnesses would know that even if they choose a foil (in target-absent lineup), they may have an opportunity to select the culprit later. If witnesses are allowed to choose more than one person then many of them do so (Levi, 1998; Lindsay, Nosworthy, Martin, & Martynuck, 1994).

However, to date only a limited number of studies have involved multiple-choice sequential lineups. For instance, Levi (2006) investigated the effect of modified sequential lineups (MSL), which are larger (containing 12, 24 or 48 members per lineup) than typical lineups, and of allowing multiple choices. He found that MSL lineups produced more identifications than typical sequential lineups (34% vs 10%) and that multiple choices increased identifications compared to sequential lineups. Thus, on the one hand allowing multiple choices could result in more identifications. However, this may also increase the number of choices, resulting in more false identifications.

Wells and Pozzulo (2006) studied eyewitnesses' accuracy for a two-culprit crime. One hundred and fifty participants saw a videotaped theft involving two culprits. Their study used two traditional (simultaneous and sequential) lineups and a newly developed procedure called 'the two-person serial lineup' (where lineup members were presented in pairs but not the two suspects in the same pair). When the culprits were present in the lineup, then identification accuracy did not vary as a function of lineup procedure. However, when the culprits were absent there was a trend for the two-person serial lineup to produce higher correct rejections than the simultaneous or sequential lineups.

The effects of having multiple targets of different ethnicities on recognition accuracy seems to have been studied in only one study. Rachlew (2007) found the cross-race effect in Norwegian police officers and civilian police staff (N = 120),

who were exposed to two male targets, one of Pakistani and one of Norwegian origin. The own-race target was recognized 57.7% of cases opposed to 48.5% for other-race target, and only 21.6% of the participants provided correct responses for both targets. However, the participants who saw recognition sets as photospreads (compared to live or videotaped recognition set) scored significantly higher when asked to identify the own-race target.

6.2.3. Cross-race effect in multiple presentations

When several perpetrators from different ethnicities commit a crime then not only the method of lineup presentation will influence the identification of them but also their ethnicity could be a factor. Previous studies have tended to support the ´cross-race` effect (or ´own-race` bias) where people recognise their own-race faces better than other-race faces (in field and laboratory-based studies, see Chapter 5 for an overview). Several psychological explanations (also see Chapter 5) have been suggested to explain the cross-race effect such as social attitudes (in early publications) and the possible influence of interracial contact and the notion of perceptual learning (in the more recent literature - Meissner & Brigham, 2001).

In Study 4 in this thesis it was found that in Estonia own-race faces were *not* more correctly identified than other-race faces, which does not support the notion that people are better at recognizing own-race faces (Meissner & Brigham, 2001; Pezdek, et al., 2003; Sporer, 2001a; Wright, et al., 2001). This finding supported more Levin's ´other-race classification advantage` (1996) (see Chapter 5). However, for correct rejections the own-race preference was present which supports Sporer's (2001a) in-group/out-group model of face processing (i.e. when in-group faces are processed more holistically whereas out-group faces more featurally). This may have a larger effect on correct rejections than correct identifications, especially concerning the accuracy of other-race faces.

Previous research indicates that in a target-present dataset correctly rejecting faces is more difficult than correctly identifying a face, especially for children (see Study 2; Pozzulo & Lindsay, 1998). The findings of Study 4 (in Chapter 5) indicate also

that the mode of processing could be different between correctly (i) recognizing and (ii) rejecting own- and other-race faces.

6.2.4. Gender differences in cross-race face recognition.

In general, females have found to recognise faces better than males do (Shapiro & Penrod, 1986; de Frias, Nilsson, & Herlitz, 2006; Schretlen, Pearlson, Anthony & Yates, 2005). However, the own-gender bias effect (that the males recognize better the male faces and females male faces) has been largely found for females but not for males. Rehnman (2007) found that females recognized better female faces, and that this effect was present in recognition of own- and other-races faces. However, for male faces, some studies report gender differences favouring females, some no gender differences, and other that males outperform females (e.g. Ellis et al., 1973; Feinman & Entwistle, 1976; Lewin & Herlitz, 2002; McKelvie, 1981; 1993; Wright & Sladden, 2003).

In Shapiro and Penrod's (1986) meta-analysis females were more likely than males to make correct identifications but were also more likely to make false positives (see also de Frias et al., 2006; Schretlen, et al., 2005). Valentine, et al. (2003) found that witness gender had a significant effect only on the number of mistaken identifications of foils in that while almost identical proportions of males and females identified the suspect (41% vs 40%), males were less likely than females to identify a foil (17% vs 28%, respectively) and correspondingly more likely to make no identification (41% vs 31%). Thus, females were more likely to choose from the lineup but no more likely to identify the suspect. As there are few studies which have examined gender differences in cross-race face recognition, the present study investigates this notion more closely.

6.2.5. Present study

It is proposed that there could be a shift in processing between (i) correct identification and (ii) correct rejection, where the correct identification of several ethnicities is different from their correct rejection. Therefore, it is examined whether

the trends from Study 4 will be present (i.e. the differences between own- and other-race face recognition emerge more concerning correct rejections).

There exists limited research on the effect of multiple choices in a sequential presentation. The participants in Study 4 made more false alarms compared to the false alarm rates typically found with usual sequential lineups. Thus, seeing different race faces together in random order (as in Study 4) could be a more difficult task than seeing several different mini-lineups where for each lineup only one race is present. Therefore, in the present study the new presentation method introduced in Study 4 is compared with presenting single-race line-ups one after another to examine whether this method can increase the accuracy rates in multi-ethnicity face recognition. It is hypothesized that participants more correctly recognize faces using mini-lineups.

Finally, gender effects concerning different race face recognition and presentation methods are investigated. It is hypothesized that gender has an effect on face recognition accuracy.

6.3. METHOD

6.3.1. Participants

The participants were young adults (N = 225, with the average age of 17 years, range 15 to 19 years, 97 boys and 128 girls) who were recruited by notices distributed in Tallinn and Otepää, Estonia. All the participants spoke Estonian as their primary language and were Caucasian.

6.3.2. Design

A 3 (type of lineups; ´decide all` vs ´decide some` vs ´sequential lineups`) × 2 (own-race (Caucasian) vs other-race faces (Afro-American, Latino, and Turkish faces)) × 2 (type of presentation; target-present vs target-absent) × 2 (gender; males, females) mixed design was used.

6.3.3. Procedure

The procedure was similar to Study 4 (in Chapter 5). Young adults were tested in small groups and within these were allocated randomly into different test groups. Firstly, the male experimenter entered a classroom where participants were together with their teacher and introduced himself. Each participant was shown a set of four male stimulus faces (on each stimulus face sheet there were four faces of different ethnicities, Caucasian, Afro-American, Turkish, and Latino, see Appendix 6 for targets and target replacements). There were four different stimulus faces sheets used in this study (see Appendix 7). These stimulus faces were shown to participants for about 30 seconds. Then the participants completed a short personal questionnaire in order to create a delay between face presentation and recognition.

There were two different face recognition sets for all the conditions (see below). Each of the face recognition sets employed had two practice and twenty four test faces which were shown sequentially for 15 second each (see Appendix 8). Two of the originally seen four stimulus faces (i.e. of different ethnicities) were present among these 24 faces (this option is referred as a target-present dataset), and the other two stimulus faces were not (target-absent dataset). There were six faces of each ethnicity present in the recognition set.

Face recognition sets were divided into three different test groups that are here called 'sequential lineups', 'decide all' and 'decide some'. It is important to note that in this study the sequential lineups were conducted in a way they are in UK (Home Office, 2002), where the participants will see all the faces and then have to make a decision (compared to the meaning of 'typical sequential lineup' in the psychological literature where the lineup is stopped when the participant makes a decision). Therefore, UK identification system is used in the present study and is called 'sequential lineup'.

In the *first* test group *sequential lineups* containing four different but consecutive mini-lineups were presented to participants (i.e. for each ethnicity a separate mini-lineup was presented). Two of the four mini-lineups were target-present and two were target-absent. The participants were told that the targets might or might not be

in the lineup and therefore they could also choose an option 'not there'. The participants saw one mini-lineup at a time and then had to make their decision. The order of presenting the different race mini-lineups was randomized.

In the *second* test group (*decide all*) the participants were shown sequentially 24 different race faces (same faces as for the 'sequential lineups', above) and they had to decide now for each of the face whether the face was or was not one of the four faces they had seen earlier (i.e. covert decision). Thus they had to make a decision about *each* of the 24 faces in the recognition set.

In the *third* test group (*decide some*) the participants were also shown sequentially 24 different race faces and they had to make an overt decision only for those faces which they thought were identical to the four faces they had seen earlier. For both 'decide all' and 'decide some' test groups different race targets and foils were randomly placed in all positions across the test array (i.e. the participants saw a mixed set of own- and other-race faces).

When interpreting the results, Caucasian faces are considered to be as 'own-race' faces in the present study and Afro-American, Latino, and Turkish faces are grouped as 'other-race' faces.

6.3.4. Test faces

A similar procedure to that in Study 4 was used. The foils in the test set of faces were selected on the basis of similarity ratings using the procedure of Pigott and Brigham (1985). The test set comprised 24 black-and-white headshots of the face. Different race faces were presented randomly in all positions across the recognition set.

6.4. RESULTS

The proportions of correct and false identifications in the target-present datasets and the proportions of correct rejections in the target-absent datasets were computed. First will be presented information concerning the effects on own- and other-race

face recognition. Then gender effects on face recognition are studied. Finally, the comparison of the three different presentation methods will be examined.

Similarly to Study 4, logistic regression analysis was conducted. Also, other-race faces (Afro-American, Turkish, and Latino) were analyzed together to be able to conduct logistic regression. It should be mentioned that for the 'decide all' and 'decide some' presentation methods it was easier for participants to make multiple choices than for 'sequential lineups' because of the nature of presentation (participants saw four different mini-lineups and after each lineup they had to make a decision about whether the target was present or not). Therefore, it should be noted that the proportion of false identifications can be higher for 'decide all' and 'decide some' presentation methods than for 'sequential lineups'.

6.4.1. Effects of different races.

First, the descriptive data in own- and other-race face recognition are presented. Table 6.1 presents the percentages and frequencies of participants choosing the target correctly (correct identification) or choosing foils (false identification) for the target-present dataset; and the percentage of correct rejections for the target-absent dataset.

Table 6.1. *Percentages and Frequencies of Correct Identifications, False Identifications, and Correct Rejections of Own- and Other-Race Faces.*

	Own-race faces	*N*	Other-race faces	*N*
Correct identifications	46%	116	63%	225
False identification	5%	116	50%	225
Correct rejections	68%	116	19%	225

Logistic regression revealed that own-race faces were more often correctly rejected than other-race faces B = .94 (.31), B(Exp) = 2.56, p = .046 (R^2 = .04 (Cox & Snell), .05 (Nagelkerke), model $\chi^2(1)$ = 4.13, p = .042). There were no differences in correct and false identification.

6.4.2. Gender effects in different race face recognition

Now the descriptive data of gender effects in own- and other-race face recognition is presented. Table 6.2 presents the percentages of males and females regarding correct and false identifications, and correct rejections of own- and other-race faces.

Table 6.2. *Gender Effects and Frequencies in Correct Identifications, False Identifications, and Correct Rejections.*

	Male	*N*	Female	*N*
Own-race				
Correct identifications	65%	65	39%	51
False identification	6%	65	4%	51
Correct rejections	63%	49	72%	67
Other-race				
Correct identifications	43%	97	67%	128
False identification	57%	97	45%	128
Correct rejections	19%	128	21%	97

The differences between males and females in own- and other-race face recognition are now investigated. Logistic regression revealed that males identified more correctly own-races faces than did females B = 1.04 (.39), B(Exp) = .35, $p < .01$ (R^2 = .06 (Cox & Snell), .08 (Nagelkerke), model $\chi^2(1) = 7.48$, $p < .01$). However, females identified more correctly than males other-race faces B = -1.19 (.34), B(Exp) = .31, $p < .01$ (R^2 = .07 (Cox & Snell), .09 (Nagelkerke), model $\chi^2(1) = 8.64$, $p < .01$)

6.4.3. Effect of different presentation methods on face recognition

Now the effects of the different presentation methods are analyzed (see Table 6.3). There were no effects of presentation methods on the recognition of own-race faces. However, there were effects for other-race faces, namely other-race faces were more falsely identified with 'decide all' and 'decide some' presentation methods than with 'sequential lineups B = .90 (.17), B(Exp) = 2.46, $p < .001$ (R^2 = .13 (Cox & Snell), .17 (Nagelkerke), model $\chi^2(1) = 30.52$, $p < .001$). A similar pattern was

present for correct rejections B = .69 (.21), B(Exp) = 2.00, $p < .001$ ($R^2 = .05$ (Cox & Snell), .08 (Nagelkerke), model $\chi^2(1) = 11.57, p < .001$).

Table 6.3. *Percentages and Frequencies in Correct Identifications, False Identifications, and Correct Rejections Across Different Presentation Methods.*

	Sequential lineups	Decide all	Decide some
Own-race			
Correct identifications	50%	40%	52%
N	40	45	31
False identification	5%	7%	3%
N	40	45	31
Correct rejections	78%	62%	60%
N	42	41	28
Other-race			
Correct identifications	61%	70%	54%
N	80	86	59
False identification	26%	69%	54%
N	80	86	59
Correct rejections	34%	12%	10%
N	80	86	59

6.4.4. Effect of gender on different presentation methods

The effects of gender on different presentation methods are now examined (see Table 6.4). In own-race correct identification males performed better than females in all presentation methods B = 1.15 (.40), B(Exp) = .32, $p < .01$ ($R^2 = .08$ (Cox & Snell), .11 (Nagelkerke), model $\chi^2(2) = 9.57$, $p < .01$). Both males and females recognized (i) other-race faces less falsely and (ii) rejected other-race faces more correctly using 'sequential lineups' than with using 'decide all' and 'decide some' presentation methods B = .88 (.17), B(Exp) = 2.41, $p < .001$ ($R^2 = .13$ (Cox & Snell), .18 (Nagelkerke), model $\chi^2(2) = 31.31, p < .001$) and B = .75 (.22), B(Exp) = 2.11, $p < .001$ ($R^2 = .08$ (Cox & Snell), .12 (Nagelkerke), model $\chi^2(2) = 12.73$, $p < .002$)

Table 6.4. *Gender Effects and Frequencies in Correct Identifications, False Identifications, and Correct Rejections across Different Presentation Methods.*

	Male			Female		
	SL	DA	DS	SL	DA	DS
Own-race						
Correct identifications	66	50	70	43	30	29
N	12	22	17	28	23	14
False identification	0	5	6	7	9	0
N	12	22	17	28	23	14
Correct rejections	82	65	50	77	62	75
N	11	20	16	31	21	12
Other-race						
Correct identifications	55	69	42	64	76	69
N	22	42	33	58	44	26
False identification	27	67	64	27	70	42
N	22	42	33	58	44	26
Correct rejections	45	17	9	29	9	12
N	22	42	33	58	44	26

Note. SL – sequential lineups; DA – decide all; DS – decide some.

6.5. DISCUSSION

In the present study different-race face recognition was examined further. The effects of different presentation methods were investigated as well as gender effects. Three main findings emerged. First, there was a difference in own- and other-race face recognition indicating own-race preference in correct rejection. Secondly, significant effects emerged regarding different target presentation methods where sequential lineups had an advantage over other multi-race face recognition sets concerning other-race face recognition. Thirdly, gender differences in face recognition emerged where males more correctly identified own-race faces and females more correctly other-race faces. These findings will be now discussed in closer detail.

6.5.1. Effects of different races.

Own-race faces were more often correctly rejected than other-race faces which supports the findings from Study 4. However, as in Study 4, the present study also found that Estonian people are *not* better at recognizing own-race faces in terms of correct identifications. The findings of this study therefore support Levin's 'other-race classification advantage' (1996) and contradict previous studies where people were found to be more accurate for own-race face identifications (Meissner & Brigham, 2001; Sporer, 2001a; Wright, Boyd, & Tredoux, 2001; Rachlew, 2007).

This rather surprising outcome could be due to using multiple choice presentation methods in the present study which allowed witnesses to choose more than once from sequential lineups. For example, Levi (2006) found that his MSL lineups produced more identifications than typical sequential lineups. However, as he concluded, while allowing multiple choices could result in more identification, but at the same time this may also increase the number of choices, resulting in more false identifications. This was not so apparent for own-race faces but emerged in a larger scale for other-race faces.

Another factor influencing these results could be that participants saw (similarly to Study 4) recognition sets as photographs. As Rachlew (2007) noted, participants who saw recognition set on photospreads scored significantly higher in correct identification compared to live or videotaped presentation. Therefore, the results could differ when presentation mode would be different (although Study 1 in this thesis demonstrated that at least for children the presentation method was not influencing person identification ability).

6.5.2. Effect of different presentation methods on face recognition

In the present study the faces in recognition set were shown sequentially and participants had to make a decision for each face whether or not it was the target. This presentation method is supposed to diminish the effect of using a relative judgment strategy and rather urge witnesses to use 'absolute judgment strategy' which will diminish false alarms when target is absent (Lindsay & Wells, 1985).

It was found that other-race faces were more falsely identified with 'decide all' and 'decide some' presentation methods than with 'sequential lineups'. A similar pattern was present for correct rejections (i.e. 'sequential lineups' increased the number of correct rejections). Surprisingly, no effects of presentation methods on the recognition performance of own-race faces were present. Wells and Pozzulo (2006) found that their two-person serial lineup produced higher proportions of correct rejections than the simultaneous or sequential lineups when the culprits were absent. The findings of the present study support the notion that 'sequential lineups' have an advantage in reducing correct rejections in other-race faces when target is absent (compared to when different race faces are presented in one large recognition set).

Previous studies have tended to support the 'cross-race' effect (see Chapter 5 for an overview). As in Study 4 it was found that own-race faces were *less* correctly identified than other-race faces (although this finding was not statistically significant) which does not support the notion that people are better at recognizing own-race faces (Meissner & Brigham, 2001; Sporer, 2001a) and supports Levin's 'other-race classification advantage' (1996). However, as in Study 4, an own-race preference effect for correct rejections was present which supports Sporer's (2001a) in-group/out-group model of face processing.

Alternative explanations to why other-race faces were more correctly identified could be that participants used a more liberal criterion with other-race faces than with own-race faces which resulted in so-called 'biased' identification of other-race faces (according to Doyle, 2001; Meissner & Brigham; 2001). This means that participants were more prone to choose (as they had multiple choices) and with multiple choices some of the choices were accurate and some not. If witnesses would behave like this then it could be a serious problem.

Therefore, the findings of the present study indicate (as Study 4) that the mode of processing could be different between correctly (i) recognizing and (ii) rejecting own- and other-race faces. Wells and Luus's (1990) approach is therefore supported where correct identifications could largely be the result cognitive processes whereas correct rejections could also involve social as well as cognitive factors (which

confirms than even for adults correctly rejecting a lineup when the target is absent is more difficult task than correctly identifying when the target is present).

6.5.3. Gender effects

In general, females recognise faces better than males do (Shapiro & Penrod, 1986; de Frias et al., 2006; Schretlen, et al., 2005). In the present study males were better at identifying own-race faces and females better at identifying other-race faces. It is known that the own-gender bias effect has been found for females and not for males (Rehnman, 2007). However, for male faces the findings are still ambiguous, namely some studies (like the present study) report gender differences favouring males, some no gender differences, and other that females outperform males (e.g. Ellis et al., 1973; Feinman & Entwisle, 1976; Lewin & Herlitz, 2002; McKelvie, 1981; 1993; Wright & Sladden, 2003). It could be that even for male targets, females demonstrate less 'variability' in (different race) face recognition performance (i.e. they recognise male or female face equally well) whereas males' performance are more influenced from the targets' gender (i.e. they are better with own-gender targets).

When the differences between males and females regarding different presentation methods were analyzed, then in own-race correct identification males performed better than females in all presentation methods and both males and females recognized (i) other-race faces less falsely and (ii) rejected other-race faces more correctly using 'sequential lineups' than with using 'decide all' and 'decide some' presentation methods. This demonstrated that 'sequential lineups' help to reduce false identifications and incorrect rejections. Overall, it can be concluded that gender had an effect on correct identifications but not on false identifications and correct rejections.

6.6. CONCLUSION

The findings of the present study indicate that when a multi-perpetrator crime takes place, then identifying the culprits can be difficult and will depend on different aspects of face recognition such as whether targets are from own- or other-race and

whether a target is present or absent. Some trends in Study 4 were confirmed where difficulties arise concerning the recognition of other race faces. It was confirmed that there appears to be shift in processing between (i) correct identification and (ii) rejection. The correct identification rates of different race faces can be very different compared to correct rejection rates.

The weakness of Study 4 was resolved in this study where participants had to make a decision for every face separately which is similar to ´choose all` condition in this study (although they could use the ´don´t know` option in Study 4). However, this did not have a large effect on recognition accuracy, even when participants had to make an ´overt` decision when they made a decision only for those faces they thought were similar to those of targets (´decide some` presentation method). Therefore, ´decide some` and ´decide all` sequential presentations in comparison to ´sequential lineups` will influence the results towards of having higher rate of false identifications and incorrect rejections.

From the present findings it is suggested that the culprits should be presented in different lineups than mixed into one sequential set. This is easier to conduct in police settings. However, for another professional group, namely immigration officers / border guards, correctly recognising people from various ethnicities may well remain a necessary but difficult task (e.g., because they see one person at a time and have to compare the person's face with the picture in the passport or with those previously encoded faces who are in their ´search list`). As this study demonstrated, when participants saw different race faces in a same large sequential presentation then mistaken identifications started to increase. Therefore, seeing persons one at a time and have to make a decision for each of the person is more difficult than making a correct decision in ´sequential lineups`. More research should be conducted in this field, for example how to improve border guards' or immigration officers performance' (especially now in the age of terrorism). More research is also needed to find a balance between correct and mistaken identifications of different race faces.

7. CHAPTER 7: OVERVIEW OF TRYING TO IMPROVE YOUNG PERSONS' WITNESS PERFORMANCE.

7.1. INTRODUCTION

At the beginning of this thesis some different ways of improving child and young adult eyewitness' testimony were discussed. The results of the present thesis can be summarized that it is very challenging to improve young persons' eyewitness testimony. In this chapter, the findings of the five studies will be reviewed and compared with those of prior research. Also methodological shortcoming of the current thesis will be discussed. Finally, implications for future research and an evaluation of the contribution of the present thesis will be presented. The contribution of this thesis to theory is written in bold.

7.2. SUMMARY OF EMPIRICAL CHAPTERS

The five studies presented in the current thesis involved (a) a meta-analysis of child eyewitness studies, (b) Estonian police officers' questioning styles with children, (c) using a person as a standard in young adults' person descriptions, (d) Estonian children's and young adults' face recognition ability, and (e) the comparison of three sequential presentation methods in Estonian young adults' face recognition ability.

7.2.1. Study 1.

Results. In the meta-analysis there were no effects of child age on correct identifications and false alarms, although children made more false alarms than adults. **Children's identifications were more correct when there were fewer members in the lineup. Delay did not have effect on children's correct identifications and false alarms.** For showups, children's correct identification rate was higher and false alarm rate lower than for simultaneous or sequential lineups. There were no effects of the mode of lineup presentation (live, photo or videospreads) for children.

Relationship with previous work. Previous studies have shown that children's identification accuracy increases with age (Pozzulo & Lindsay, 1998) which was confirmed also by the present meta-analysis. Children's identification abilities are similar to those of adults' in correctly identifying persons from TP lineups but for TA lineups their performance is poorer and they make more false identifications (Parker & Ryan, 1993; Gross & Hayne, 1996; Dekle et al., 1996). In this meta-analysis the error rate for sequential lineups in children was higher than for simultaneous lineups or showups which confirms previous findings (Pozzulo & Lindsay, 1998). Therefore, the previous research is supported where sequential lineups increase adults' but not children's correct rejections (Lindsay et al., 1997).

Pozzulo and Lindsay (1998), and also Dekle et al. (1996) found differences between adults and children in correct identifications. In the present meta-analysis younger children's correct identification rate was lower than older children's and adults' but no statistical difference between adults and children (all ages combined) emerged (see also Parker & Ryan, 1993; Gross & Hayne, 1996; King & Yuille, 1986). On one hand it could be due to the fact that age differences between children and adults are diminishing concerning correct identifications and on the other hand this could be due to methods of statistical analysis (e.g., random effects regression analysis were used which was more stricter than fixed effects regression analysis in Pozzulo and Lindsay's meta-analysis). Some previous studies have found false alarms to decrease with age (Chance et al., 1982; King & Yuille, 1986; Davies et al., 1988). In the present meta-analysis there were no effects of child age on false alarms (for TA lineups), although children made more false alarms than adults.

Concerning lineup types, children made more false alarms than adults for all types of lineups. In the TA condition, showups produced a higher level of correct rejections and false identifications (which is similar to Steblay et al., 2003 and Dekle et al., 1996). For children and adults there were no effects of the mode of target presentation (i.e. whether target was presented either live, via video or via photographs). Lindsay and Harvie (1998) found similar results but only for TP lineups whereas Lindsay and Pozzulo (1998) found that children correctly rejected TA lineups with a live target presentation less often than adults did. Shapiro and Penrod (1986) in their meta-analysis stated that when the event was presented in a

more realistic way (either live or on videotape) then people remembered the target more accurately than when viewing it via photographs (see also Rachlew, 2007). Therefore, it should be considered that presenting people via photographs can have an effect on performance compared to other presentation methods.

Similarly to previous research, delay did not have any effect on children's performance in TP and in TA lineups (Goodman, Bottoms, & Schwartz-Kenney 1991; Peters, 1987, 1991). A delay of up to one month could be not long enough (compared to the delays in real life).

7.2.2. Study 2.

Results. A pattern of long interviewer questions and short child answers was often apparent. **During interviews, the proportion of direct questions was found to decrease over time and the proportion of option-posing and suggestive questions to increase.** Longer answers were provided in response to general or central invitations, whereas option-posing or closed questions produced less information. **Children reported more judicially significant details in response to invitations than to option-posing and suggestive questions.**

Relationship with previous work. Previous studies also confirm that open memory prompts produce longer and more accurate responses than do closed prompts (Hershkowitz, Lamb, Sternberg, & Esplin, 1997; Lamb, Hershkowitz, Sternberg, Esplin, et al., 1997). Similarly to Study 2 it was found in previous research that interviewers tend to use less open questions and more closed questions throughout interviews (Sternberg, et al., 2001; Davies, et al., 2000; Lamb, Orbach, Sternberg et al., 2002; Westcott & Kynan, 2006). Also, free recall prompts produced, as expected, more information than questions (Lamb, Hershkowitz, Sternberg, Esplin, et al., 1997; Sternberg et al., 1996; Sternberg, et al., 2001; Korkman, et al., 2006).

However, the results of present study contradict those of Cederborg et al. (2000) who found in Sweden that option-posing and suggestive questions were often introduced early in interviews. Innovatively, children reported more judicially significant details in response to invitations than to option-posing and suggestive

questions. Thus, similarly to other countries, investigative interviews in Estonia tend to rely also more on directive and option-posing questions.

7.2.3. Study 3.

Results. **There was no overall effect of the use of a standard but the use of a 'typical' male standard increased the amount of information about characteristics that would otherwise not be remembered well.** The findings supported to some extent the 'own-anchor' effect in that male observers who were physically similar to male target were more accurate (and persons who were dissimilar were either less accurate / made more mistakes).

Relationship with previous work. A novel approach of how to increase the amount of information about a once-seen person, namely by using a 'standard' in witnesses' visual field at the time of giving the person description was examined. Using a standard could assist witnesses to recall more information concerning characteristics which are not otherwise remembered well such as inner face features and figure (Kuehn, 1974; Ellis, 1990; Sporer, 1992b; Lindsay, Martin, & Weber, 1994; Van Koppen & Lochun, 1997; Pozzulo & Warren, 2003). However, Kask, et al. (2007) found that provision of 'standard' did not help children overall to recall more information about a once-seen person and in Study 3 this effect emerged also for adults.

There were no effects of gender on young adults' free descriptions which contradicts the findings of Powers, et al. (1979) and Yarmey (1993) who found females to outperform males in some target's characteristics. As ethnic differences were not present, the results of this study confirm Fallshore and Schooler's (1995) findings which state that there was no relationship between description accuracy and identification performance for 'other-race' faces. Similarly to Sporer (1992b) and Lindsay et al. (1994) this study found that young adults' free recall was emphasizing more on characteristics such as hair, face, and clothes.

As in Flin and Shepherd's research (1986), in this study 'own-anchor' effect was found only for some characteristics concerning male observers describing a male target. Similarly to Biernat et al. (1991) an own-gender effect was present in peoples' height and weight estimations (e.g. participants who were similar to the target were more accurate in estimating target's characteristics than those who were less similar to the target).

As a novel finding, typicality of standard effect emerged (when the standard was similar to target then it helped to recall more details than when standard was different from the target). The findings indicated that as the nature of effect is not very clear yet, more research should be conducted to investigate different standards' possible influence on assisting person descriptions.

7.2.4. Study 4.

Results. For own-race faces participants made more correct rejections than for other-race faces. **For other-race faces correct recognition decreased when similar foils were presented before targets in the recognition set but this did not occur for own-race faces.**

Relationship with previous work. In this study 'cross-race' effect was found similarly to previous research (Meissner & Brigham, 2001; Sporer, 2001a; Wright, et al., 2001; Shapiro & Penrod, 1986). An own-race effect for children was also present which confirms other recent findings (Pezdek, et al., 2003; Corenblum & Meissner, 2006; Sangrigoli, et al., 2005; Goodman et al., 2007; Sporer, et al., 2007). Interestingly, differences between children and adults did not occur for correct identifications or correct rejections for both own- and other-race faces.

The participants more correctly recognized some other-race faces than own-race faces. 'Attitude' theory may not explain these findings (Berger, 1969; Galper, 1973) and a more complex explanation may well be needed. The results found support to Levin's (1996) 'other-race classification advantage' theory according to which people are slower at classifying the race of own-race faces compared to other-race

faces. Also, it could be that participants were more willing to guess concerning other-race faces (Doyle, 2001) and this produced larger amount of correct identifications in this study. If this is valid then Meissner and Brigham's (2001) conclusions are confirmed where witnesses could have a more liberal criterion with other-race faces than with own-race faces and own-race preference bias emerge more in false than in correct identifications.

In this study, to increase ecological validity regarding the initial seeing of several targets of different races the participants had the opportunity to make multiple choices in a multi-race sequential recognition set. **This new type of target presentation demonstrated similar results to those conducted using traditional lineups.** The high percentage of correct identifications could be due to the method used to conduct this study (e.g. photospreads). For example, Rachlew (2007) found that participants scored significantly higher when recognition set was presented on photospreads (and not via live or videotaped presentation). Shapiro and Penrod's (1986) meta-analysis confirms Rachlew's results.

Similarly to this study, Clark and Davey (2005) demonstrated that the most similar-looking foil was chosen at a higher rate if he appeared in the order late rather than earlier. This effect emerged especially for children. Therefore, the attention to the position of target and foils should be more addressed when conducting lineups either in practice or in scientific purposes. This finding support Slone et al.'s (2000) notion which stated that the own-race effect may occur due to considering new faces of different ethnicities as familiar/similar (i.e. ´seen before`) compared to the accurate recognition of own-race faces.

There were no effects of race regarding the total number of descriptors provided. For the various descriptor items, only some differences emerged between different race faces which are in line with previous findings (Fallshore & Schooler, 1995; Meissner & Brigham, 2001; Sporer, 2001a, 2001b).

Finally, some but not all aspects of the ´cross-race` effect was also present in a less ´westernised` country such as in Estonia. However, the cross-race effect should be examined further because it could still not be universal and could apply more to

Caucasian than other-ethnicity participants (see also Anthony et al., 1992; Chiroro & Valentine, 1995).

7.2.5. Study 5.

Results. Own-race faces were more often correctly rejected than other-race faces. Males identified more correctly own-races faces and females identified more correctly other-race faces. **There were no effects of presentation methods on the recognition of own-race faces but the differences emerged for other-race faces in favour of sequential presentation methods.**

Relationship with previous work. Similarly to Study 4, own-race faces were more correctly rejected than other-race faces and likewise own-race faces were less correctly identified than other-race faces (although not statistically significantly). The results concerning correct identification contradict previous research where people were found to be more accurate in own-race face identifications (Meissner & Brigham, 2001; Sporer, 2001a; Wright, Boyd, & Tredoux, 2001; Rachlew, 2007) and could be taken to support therefore Levin's 'other-race classification advantage' (1996).

Previous research indicates also that in a target-present dataset correctly rejecting faces is more difficult than correctly identifying a face, especially for children (see Study 1; Pozzulo & Lindsay, 1998). Therefore, Sporer's (2001a) in-group/out-group model of face processing is supported according to which in-group and out-group faces are processed differently (and this could also have an effect to differences between correct identifications and correct rejections of own- and other-race faces). Namely, the encoding strategy could influence also person identification in that own-race faces are more accurately recognised than other-race faces (and when the target is present then identification relies also more on holistic strategy whereas when target is absent then featural strategy emerges which could explain why there is a large gap between (especially children's) correct identifications and rejections).

It was found that 'sequential lineups' had an advantage in increasing correct rejections when the target was absent (compared to when randomised different race

faces were presented in one large recognition set such as 'decide all' or 'decide some' in this study). One possible explanation of this finding could be the availability to multiple choices. Levi (2006) stated that multiple choices result in more identification and also in more false identifications (especially for other-race faces). Similarly to Study 4, the factor that participants saw recognition sets as photographs should be considered (Rachlew, 2007). [Although Study 1 in this thesis demonstrated that presentation method was not influencing person identification accuracy (for children).]

Females have found to recognise faces better than males (Shapiro & Penrod, 1986; de Frias, Nilsson, & Herlitz, 2006; Schretlen, et al., 2005) although in Study 5 mixed results concerning gender effect was found. The own-gender bias effect has been found to exist in females (Rehnman, 2007), however for male faces some studies (like the present study) report gender differences favouring males, some demonstrate no gender differences, or that females outperform males (e.g. Ellis et al., 1973; Feinman & Entwisle, 1976; Lewin & Herlitz, 2002; McKelvie, 1981; 1993; Wright & Sladden, 2003). As stated in Study 5, females' face recognition performance could be more 'universal' as they seem to recognise male or female face equally well whereas males' performance are more influenced by the own-gender effect.

Wells and Luus (1990) suggested that correct identifications are largely the result cognitive processes whereas correct rejections also involve social as well as cognitive factors. The findings of Studies 4 and 5 indicate that the correct identification of different race faces is different from correct rejection (similarly to Study 1 in this thesis and to Pozzulo & Lindsay, 1998) and a shift in processing exists between (i) correct identification and (ii) rejection.

Carey and Diamond (1977) suggested that before the age of eight (until the age of six according to Tanaka et al., 1998 and Pellicano & Rhodes, 2003), children recognize faces featurally and then switch to more holistic recognition strategy. They propose that own-race faces would be processed holistically whereas other-race faces featurally which results in own-race superiority effect. However, it is still unclear *when* the shift from holistic to featural encoding strategy emerges. As stated

above, Sporer's (2001a) in-group/out-group model of face processing could be applied to those differences in children's face recognition.

7.3. METHODOLOGICAL AND STASTISTICAL ISSUES

As with most applied research, designs and procedures are often not perfect. Now some of the strengths and weaknesses of different studies presented in this thesis are discussed.

A strength of Study 1 is that in the regression analysis a random effects methods was used instead of fixed effect model like in Pozzulo and Lindsay's (1998) meta-analysis. This allowed more sophisticated analysis to be conducted (see Mundlak & Yahav, 1981).

Very few researchers in the forensic area ever present in their publications the nature of data gathering when conducting research with children. Conducting eyewitness research with children is a very challenging task. Verbal transcripts have to be translated into quantitative memory units and coding schemas can be difficult for coders to apply into practice and in some cases it is even impossible to decide the correct category. This is especially valid in decision-making process about whether certain aspects in child's recollection were relevant or not (see Study 2). Also, repetitions in children's accounts are very difficult to code and differences across participants often result in large variations within the raw data.

For example, in Study 2 when specific or open-ended questions were asked there was large variation among the children. Some children did volunteer a great amount of detail while others failed to report even the major aspects. In general, although children did respond more or less correctly to the questions, their free recall itself in this study was generally vague. It has to be mentioned that the categories in this study were strict because the sample was small ($N = 17$) and therefore every mistake or disagreement between coders would have a more distinctive effect on the results than in a study with larger sample.

Also, concerning Study 2, it was found very hard to get access relevant organizations such as police forces. Usually obtaining official permission is the first

stage in this process followed by negotiations with individuals who will actually participate in the research (Horn, 1996). Several researchers in applied settings have described how they were seen as ´spies` by others in the setting (Horn, 1996; Hunt, 1984, Warren, 1988). The presence of a researcher is feared (Lee, 1993) because of the possibility that they might either reveal deviant activities or other matters. For example, when data were collected for Study 2, only transcripts conducted by officers not working there any more were provided.

In a highly structured organization such as the police, the researcher may obtain official permission to conduct a study but receive only moderate or little co-operation from those lower down the hierarchy (Dingwall, 1980). Also it might be that the participating is not being seen as voluntary but rather as compulsory by the participants (Horton & Smith, 1988). One can have formal access to the police force but not informal access. Therefore, it is essential for the participants to understand the aim of the study and feel that they can contribute to it (Horn, 1996) and they must feel that the researcher is trustworthy (Buchanan, Boddy, & McCalman, 1988).

Usually such so-called ´messy` aspects are not reported by psychologists (e.g. Hunt, 1984), preferring to present themselves as objective scientists, and their research as straightforward and unproblematic (Horn, 1996). In fact, applied research is messy, and researchers often face situations where the ´correct` response is not clear. Bryman (1988) describes the situation very well by saying that researchers ´´brought up on a diet of text books and sanitized research reports sometimes report their feelings of something being wrong with themselves when things do not go according to plan... It may be far more responsible to make prospective researchers aware of such facts in advance than to imbue them with self-doubt as their plans go awry`` (p.9).

In the present Study 5 it was also hoped to compare the different race face recognition ability of British and Estonian police officers. However, such participants were not available because the process to obtain access to the participants was excruciatingly slow and the official permission was not received (in time).

The event in Study 3 was aimed to represent the situation where witnesses might often be involved (similar to a theft or an accident). The main part of the mock situation did not have any criminal elements and thus would be considered as irrelevant by some police investigators. However, with those mock situations two aspects need to be considered. Firstly, in almost all criminal cases when some event occurs it is relatively brief, i.e. participants will see the target only during a limited amount of time (Ceci & Bruck, 1995). Secondly, most of the research involves episodes similar to the crimes in real life (e.g. accident or theft, cf. Flin, Boon, Bull, & Knox, 1992). Therefore, the mock situation in Study 3 was somewhat similar to what can happen with witnesses in real life.

In Study 3 some difficulties in coding emerged in estimating target's parameters (like height, weight, type of figure etc.). Very few published studies have described in details how such estimates were compared, especially in terms of how to measure accuracy (in this study these estimations were based on those used by Flin & Shepherd, 1986). As many studies do not cover these characteristics in details, then it was difficult to produce a perfect coding system for this study.

A relatively large weakness of Studies 4 and 5 was that in those studies the faces which were included in the lineup presentation sets were based on similarity to the target and not description matched. Wogalter, Malpass, and Berger (1993) noted that police officers still use match-to-suspect strategy. However, Clark and Tunnicliff (2001) proposed that a description-matched foil selection method should be used instead because the suspect-matched category is more dangerous especially in target-absent lineups where a witness could choose the next-best match instead of correctly rejecting the lineup. Therefore, the results concerning correct rejections could be influenced by the fact that faces were not selected by using match to descriptions. For example, Clark and Davey (2005) noted the "next-best lineup alternative" where most of the foil identifications in the TA lineup went to the foil photograph that appeared to be the next-best match.

In Study 4 and 5 there was also a dilemma concerning how to code the data into correct identifications, false alarms and correct rejections due to the difference of the novel presentation method from usual sequential lineups (for the data to be

comparable). For ´sequential lineups` this task was easier, namely in target-present lineup hits can be coded as correct identifications and misses as false identifications and saying ´not here` as incorrect rejections, whereas in target-absent presentation ´not here` can be coded as correct rejections and choosing someone from lineup as incorrect rejections. For the new sequential presentation methods this was achieved by hits being coded as correct identifications and misses as false identifications in target-present condition whereas in target-absent condition the categories were either correct or incorrect rejections (regarding whether they correctly rejected the faces in recognition set or not). However, in ´sequential lineups` it was difficult to make a multiple choices (and participants never did) compared to the new multiple sequential presentation methods. Therefore, some of the statistical differences between presentation methods could be due to the coding system and not presentation method. This should be considered when interpreting the results of this study.

One more weakness with Study 4 was that faces from some ethnicities were not so well represented as some other ethnicity faces (for instance, there were more Caucasian, Latino and Afro-American faces than Turkish and Chinese faces). A strength of this is that there are few prior studies with more than one ´other-race` present (Goodman et al., 2007; Sporer, et al., 2007). However, the other side of the coin is that the analyses could be done only between own- and other-race (combined) faces because (i) of the lack of representative faces for each ethnicity and (ii) the number of participants in this study (which nevertheless was fairly large). This was improved in Study 5 where all race faces had five to six foils in the face recognition set.

Overall, the present thesis did address issues of improving person description and recognition methods, children's person identification ability, and police officers' questioning methods. It is contended that in spite of the strengths and shortcomings, the validity of main findings in the present thesis in unlikely to have been reduced substantially. Yet future research needs to consider the above-mentioned criticisms and weaknesses in order to plan new experiments (for example, trying to apply a new method for recognizing multi-ethnicity faces in cases of multi-perpetrator crimes). Suggestions for future research are now presented.

7.4. FUTURE RESEARCH

Own-group bias in memory for human faces has been widely reported (Bäckman, 1991; Cross et al 1971; Meissner & Brigham, 2001; Slone et al 2000; Wright & Stroud, 2002; Anastasi & Rhodes, 2005). For example, Lindholm (2005) found that younger adults had a better recall of own-age than of other-age targets, while older adults were unaffected by the race (see also Rehnman, 2007). Own-age advantage was also reported in seven-year-old children. As there are very few studies where children's face recognition is tested on own and out-group faces, this area would be important to study both theoretically and in terms of applied procedures.

On one hand, children's development concerning face recognition is seen as continuum. It is not clear yet at which point in children's development shifts from holistic to featural encoding processing strategy is occurring. For example, Tanaka et al. (1998) noted that children recognize faces holistically by the age of six years (also see Pellicano & Rhodes, 2003). On other hand, studies have confirmed that at some point their processing abilities change very rapidly which suggests that perhaps the change is more categorical than continuous. Therefore, this should be studied further to clarify from which age on the processing of faces will change (i.e. from featural to a holistic encoding strategy).

For more precise proposals, new experiments should involve children from four to twelve years old and several comparisons should be conducted in (i) both own and other-race face recognition and (ii) in own and out-group face recognition. Such a large project could provide some insight into theory concerning when do age related changes in face recognition occur and whether there is also a shift from / to own-group recognition bias (namely that children might recognize children's faces better than adult faces).

Both Study 3 and Kask et al. (2007) have found that there was no effect of using a standard to enhance person descriptions in children and adults. It might be that children still did not completely understand what is really meant by to compare the memory about once-seen stranger with an adult standard in their visual field. However, an interesting effect was present on adults. Some adults during the testing (when standard was not used) asked experimenter what are his/her characteristics to

be able to better judge stranger's characteristics. This could give an insight that this kind of comparison is useful for *some* people. However, it is still unclear why ´use of a standard` seems not to be effective.

It might be that when a standard is very similar to the perpetrator it aids recall more, but if it is rather different then the effects might be rather negative. To study the effect of a standard more, several assumptions are made. First, in this thesis the participants did see the stranger in groups and had to answer about the characteristics also in groups. It could mean that they did not see the target very well (which, on other hand, happens also frequently in real life in that witnesses might not see the culprit very well). Secondly, they might not been very interested in responding as accurately or eagerly as when interviewed one-to-one basis. Therefore, an additional study should be conducted where participants still could see target in group but person descriptions and comparisons with standard are collected individually.

Concerning children, it could be that own-group persons would be more easily described that those from out-group (Anastasi & Rhodes, 2005) because children are short and weigh less compared to adults or even teenagers. It could add some new theoretical knowledge to conduct an experiment where children see a child as target and then have to describe him/her using another child as a standard to examine whether there is a (a) mismatch in misunderstanding the task or (b) the adult standard would not be as effective as a more own-age standard for children.

Some studies have been conducted with child witnesses to test either simultaneous or sequential lineups (see Study 1). However, there is not much research available with children using showups. There is also a lack of studies with more ecologically valid delay between seeing the stimulus and the recognition set, especially in children. In most studies the delay is from five minutes to one month which is a relatively short time concerning how long the identification process could take place in court systems. Also, several presentation methods could be compared (live vs photo vs video), especially due to the lack of data using live presentations. However, the meta-analysis in the present thesis demonstrated that presentation method is not a strong predictor compared to children's age or the type of lineups.

Another ecologically important variable, namely the effect of anxiety on children's face recognition, has largely been overlooked to date. Although issues like children's errors in accounts of traumatic injury (Peterson, 1996) or the effects of stress for a natural disaster (Bahrick, Parker, & Fivush, 1998) have been examined, only a very few studies have examined the notion concerning person identification. Some studies have found that higher stress is associated with increased recall and reduced suggestibility in children (Goodman, Aman & Hirschman, 1987; Ochsner & Zaragoza, 1988), whereas some note that heightened arousal never increased the recognition or recall accuracy and had rather detrimental effect on memory (Peters, 1991). However, Goodman, Bottoms, and Schwartz-Kenney (1991) found no effect of stress on children's accounts.

More recently, Almerigogna, Ost, Bull, and Akehurst (2007) conducted a study with eight to eleven-year-old children and they found that those children who scored highly on trait and post-interview state anxiety measures more often responded incorrectly to misleading questions. Also, pre- to post-interview changes in state anxiety were correlated with more incorrect responses to misleading questions.

One way how anxiety could be measured would be by asking children themselves after the task to rate how anxious they were during decision-making process. The other, and may be more reliable, way would be to record children's behaviour during their performance when recognizing persons from photographs or live presentations and then let adults to rate whether children seemed to be nervous. This could afterwards to be linked with children's accuracy in their performance. However, methodological issues should be considered in how the levels of stress (or anxiety) during the identification process are measured (Goodman, 1991; Peters, 1991).

Also, as there are not many studies using moving images in sequential identification procedures among children, then procedures such as Video Identification Parades Electronic Recording (VIPER) could be used. In adults it has been found that moving images reduce especially mistaken identifications in TA lineups (Valentine, Darling, & Memon, 2007). As children's performance in correctly rejecting target-

absent lineups is poor, then this could be a method which would aid them to increase accuracy in identification.

Finally, very few studies have allowed children (and adults) to say 'I don't know' when trying to identify a person from a lineup. This option seems to have a positive effect (although only two studies with such an option were available to be included into the meta-analysis). Therefore, more studies examining this effect on person identification accuracy should be conducted.

7.5. CONCLUSION

What recommendations should be given to applied practice regarding the topic of this thesis? The present thesis has shown that the topic under investigation is a difficult area to study and as past research has found, increasing the accuracy, quality and quantity of children's and young adults' performance is a challenging task. The meta-analysis demonstrated that the most important variables influencing children's performance seem to be children's age and type of lineups.

When examining Estonian investigators' questioning styles they were found to be similar to those investigators in Western Europe who have received minimal training in interviewing children. This thesis also demonstrated that the use of a standard to increase the number of correct details in young adults' person descriptions did not succeed. Therefore, further training to investigators regarding questioning styles should be conducted.

In this thesis also alternative presentation methods were introduced when recognizing several multi-ethnicity culprits in children and adults. The differences in face recognition emerged more in (i) other- than own-race face recognition and (ii) in correct rejections than in correct and false identifications. Sequential lineups were more efficient in face recognition, however, in real life we still see people one at a time such as procedures in Study 4 and 5, and therefore correctly identifying persons will not be as accurate as in using sequential lineups.

In applied settings it is suggested that firstly police officers should be more trained in (i) how to interview children and (ii) how to construct a lineup. It is important to

consider children's age when conducting a lineup or asking person descriptions. More effort on constructing a recognition set which would increase children's performance especially in target-absent cases should be done (especially in case of multiple culprits from different ethnicities).

This thesis has made an important contribution to forensic psychology as it has shown the differences and similarities between child and adult witnesses. Furthermore, it has shown that child witnesses can be accurate when their developmental needs are considered. Finally, although age plays a significant role on children's accounts regarding free recall; face recognition and person identification, findings in this thesis reveal that age differences concerning correct identifications may be decreaseing. Therefore, more attention should be directed to the similarities (and not so much differences) when between children's and adults' eyewitness performance. However, more research is still required in increasing children's performance when the target is absent.

It is clear that this thesis has raised more questions than it managed to answer. Due to the continuously increasing attention now being paid in various countries to child witnesses' performance, it is hoped that the current findings will contribute significantly to the existing body of knowledge and will encourage more in-depth examination in this important field of research.

8. REFERENCES

Agnew, S.E., Powell, M.B., & Snow, P.C. (2006). An examination of the questioning styles of police officers and caregivers when interviewing children with intellectual abilities. *Legal and Criminological Psychology, 11,* 35-53.

Aldridge, J., Lamb, M., Sternberg, K., Orbach, Y., Esplin, P., & Bowler, L. (2004). Using human figure drawing to elicit information from alleged victims of child sexual abuse. *Journal of Consulting and Clinical Psychology, 72,* 304-316.

Almerigogna, J., Ost, J., Bull, R., & Akehurst, L. (2007). A state of high anxiety: How non-supportive interviewers can increase the suggestibility of child witnesses. *Applied Cognitive Psychology, 21,* 963-974.

Anastasi, J.S. & Rhodes, M.G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review, 12,* 1043-1047.

Anthony, T., Copper, C. & Mullen, B. (1992). Cross-racial facial identification: A social cognitive integration. *Personality & Social Psychology Bulletin, 18,* 296-301.

Bäckman, L. (1991). Recognition memory across the adult life span: The role of prior knowledge. *Memory & Cognition, 19,* 63-71.

Bahrick, L.E., Parker, J.F., & Fivush, R. (1998). The effects of stress on young children's memory for a natural disaster. *Journal of Experimental Psychology: Applied, 4,* 308-331.

Baker-Ward, L., Gordon, B.N., Ornstein, P.A., Larus, D.M., & Clubb, P.A. (1993). Young children's long-time retention of a pediatric examination. *Child Development, 64,* 1519-1533.

*Beal, C.R., Schmitt, K.L., & Dekle, D.J. (1995). Eyewitness identification of children: Effects of absolute judgments, nonverbal response options, and event encoding. *Law and Human Behavior, 19,* 197-216.

Behrman, B.W. & Davey, S.L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25,* 475 – 491.

Berger, D.G. (1969). *They all look alike.* Unpublished doctoral dissertation, Vanderbilt University.

Beuscher, E. & Roebers, C.M. (2005). Does a warning help children to more accurately remember an event, to resist misleading questions, and to identify unanswerable questions? *Experimental Psychology, 52,* 232-241.

Biernat, M., Manis, M., & Nelson, T.E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology, 66,* 199-214.

Bishop, D.V.M. & Donlan, C. (2005). The role of syntax in encoding and recall of pictorial narratives: Evidence from specific language impairment. *British Journal of Developmental Psychology, 23,* 25-46.

Bothwell, R.K., Brigham, J.C. & Malpass, R.S. (1989). Cross-racial identification. *Personality & Social Psychology Bulletin, 15,* 19-25.

Brady, M.S., Poole, D.A., Warren, A.R., & Jones, H.R. (1999). Young children's responses to yes-no questions: Patterns and problems. *Applied Developmental Science, 3,* 47-57.

Brainerd, C.J., Reyna, V.F., Howe, M.L., & Kingma, J. (1990). The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development, 55,* v-93.

*Brewer, N. & Day, K. (2005). The confidence-accuracy and decision latency-

accuracy relationships in children's eyewitness identification. *Psychiatry, Psychology, & Law, 12,* 119-128.

Breznitz, Z. (2002). Asynchrony of visual-orthographic and auditory-phonological word recognition processes: An underlying factor in dyslexia. *Reading and Writing: An Interdisciplinary Journal, 15*, 15-42.

Brigham, J.C., Bennett, L.B., Meissner, C.A., & Mitchell, T.L. (2007). The influence of race on eyewitness memory. In R. Lindsay, D. Ross, J. Read, & M. Toglia (Eds.), *Handbook of eyewitness psychology* (pp. 257-281). Mahwah, NJ: Lawrence Erlbaum & Associates.

Brigham, J.C., Maass, A., Snyder, L.D., & Spaulding, K. (1982). Accuracy of eyewitness identification in a field setting. *Journal of Personality and Social Psychology*, *42,* 673-681.

Brown, C. & Lloyd-Jones, T.J. (2003). Verbal overshadowing of multiple face and car recognition: Effects of within- versus across-category verbal descriptions. *Applied Cognitive Psychology, 17,* 183-201.

Bruce, V. & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77,* 305-327.

Bruck, M., Ceci, S.J., & Francocoer, E. (1995). Anatomically detailed dolls do not facilitate preschoolers' reports of a pediatric examination involving genital touching. *Journal of Experimental Psychology: Applied, 1,* 95-109.

Bruck, M. & Melnyk, L. (2004). Individual differences in children's suggestibility: A review and synthesis. *Applied Cognitive Psychology*, *18,* 947-996.

Bruck, M., Melnyk, L., & Ceci, S.J. (2000). Draw it again Sam: The effect of drawing on children's suggestibility and source monitoring ability. *Journal of Experimental Child Psychology, 77,* 169-196.

Bryman, A. (1988). Inside accounts and social research in organizations. *In* A. Bryman (Ed.). *Doing research in organizations*. London: Routledge.

Buchanan, D., Boddy, D., & McCalman, J. (1988). Getting in, getting on, getting out and getting back. *In* A. Bryman (Ed.). *Doing research in organizations*. London: Routledge.

Buckhout, R. & Regan, S. (1988). Explorations in research on the other-race effect in face recognition. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Volume 1. Memory in everyday life* (pp. 40-46). New York: Wiley.

Bull, R. (1992). Obtaining evidence expertly: The reliability of interviews with child witnesses. *Expert Evidence, 1,* 5-12.

Bull, R. (1995). Innovative techniques for the questioning child witnesses, especially those who are young and those with learning disability. In M.S. Zaragoza, J. R. Graham, G. C. N. Hall, R. Hirschman, & Y. S. Ben-Porath, (Eds.), *Memory and testimony in the child witness* (pp. 179-194). Thousand Oaks, CA: Sage.

Bureau of Justics Statistics (2002). *Homicide trends in the United States.* Retrieved 9.12.2003 from http://www.ojp.usdoj.gov/bjs/homicide/homtmd.htm.

Burgess, M.C.R. (1997). *The cross-race effect in facial recognition: A function of expertise?* Unpublished doctoral dissertation, Florida State University.

Butler, S., Gross, J., & Hayne, H. (1995). The effect of drawings on memory performance in young children. *Developmental Psychology, 31,* 597-608.

Cain, W.J., Baker-Ward, L., & Eaton, K.L. (2005). A face in the crowd: The influences of familiarity and delay on preschoolers' recognition. *Psychology, Crime & Law, 11,* 315-327.

Campbell, R., Coleman, R., Walker, J., Benson, P.J., Wallace, S., Michelotti, J., & Baron-Cohen, S. (1999). When does the inner-face advantage in familiar face recognition arise and why? *Visual Cognition, 6,* 196-216.

Carey, S. & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195,* 312-314.

Carroo, A.W. (1986). Other race recognition: A comparison of Black American and African subjects. *Perceptual & Motor Skills, 62,* 135-138.

Cassel, W.S., Roebers, C.E.M.; & Bjorklund, D.F. (1996). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology, 61,* 116-133.

Ceci, S.J. & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113,* 403-439.

Ceci, S.J. & Bruck, M. (1995). *Jeopardy in the courtroom.* Washington, DC: American Psychological Association.

Ceci, S.J., Crossman, A.M., Gilstrap, L.L., & Scullin, M.H. (1998). Social and cognitive factors in children's testimony. In C.P.Thompson, D.J.Herrmann, J.D.Read, D.Bruce, D.G.Payne & M.P.Toglia (Eds.). *Eyewitness memory: Theoretical and applied perspectives* (p. 15-30). New Jersey: Lawrence Erlbaum Associates.

Cederborg, A.-C., Orbach, Y., Sternberg, K.J., & Lamb, M.E. (2000). Investigative interviews of child witnesses in Sweden. *Child Abuse and Neglect, 24,* 1355-1361.

Cesniene, I., & Bandzeviciene, R. (2005). Gender differences in accuracy of preschool children memory for eyewitnessed event. Paper presented in XVth European Conference on Psychology and Law. July, Vilnius.

Chae, Y. & Ceci, S.J. (2005). Individual differences in children's recall and suggestibility: The effect of intelligence, temperament, and self-perceptions. *Applied Cognitive Psychology, 19,* 383 – 407.

Chance, J.E. & Goldstein, A.G. (1979). Reliability of face recognition performance. *Bulletin of the Psychonomic Society, 14,* 115-117.

Chance, J.E. & Goldstein, A.G. (1984). Face-recognition memory: Implication for children's eyewitness testimony. *Journal of Social Issues, 40,* 69-85.

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. 153–176). Mahwah, NJ: Erlbaum.

Chance, J.E., Turner, A.L., & Goldstein, A.G. (1982). Development of differential recognition of own- and other-race faces. *Journal of Psychology, 112,* 29-37.

Chiroro, P. & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly Journal of Experimental Psychology, 48A,* 879-894.

Chung, M.-S. & Thomson, D.M. (1995). Development of face recognition. *British Journal of Psychology, 86,* 55-87.

Clark, S.E. & Davey, S.L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29,* 151 – 172.

Clark, S.E. & Tunnicliff, J.L. (2001). Selecting lineup foils in eyewitness identification experiments: experimental control and real-world simulation. *Law and Human Behavior, 25,* 199 – 216.

Clarke, C. & Milne, R. (2001). *National evaluation of the PEACE investigative interviewing course.* London: Home Office.

Clifford, B.R. (2003). Verbal overshadowing effects in adults' and children's face and voice memory. Paper presented to the International Conference for Applied Research on Memory and Cognition (SARMAC). Aberdeen, July.

Clifford, B.R. & Bull, R. (1978). *The psychology of person identification*. London: Routledge and Kegan Paul.

Corenblum, B. & Meissner, C.A. (2006). Recognition of faces of ingroup and outgroup children and adults. *Journal of Experimental Child Psychology, 93,* 187-206.

Craig, R.A., Scheibe, R., Kircher, J., Raskin, D.C., & Dodd, D. (1999). Effects of interviewer questions on children's statements of sexual abuse. *Applied Developmental Science, 3,* 77-85.

Cross, J.F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics, 10,* 393-396.

Cutler, B.L., Penrod, S.D., & Martens, T.K. (1987). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology, 72,* 629-637.

Dando, C., Wilcock, R., & Milne, R. (2008). The cognitive interview: Inexperienced police officers' perceptions of their witness/victim practices. *Legal and Criminological Psychology, 13,* 59-70.

Davies, G.M. (1996). Children's identification evidence. In S.L.Sporer, R.S.Malpass, & G.Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 233-258). New Yersey: Lawrence Erlbaum Associates

Davies, G.M. & Flin, R. (1988). The accuracy and suggestibility of child witnesses. *Issues in Criminological and Legal Psychology, 13,* 21-34.

Davies, G.M., Smith, S., & Blincoe, C. (2008). A "weapon focus" effect in children. *Psychology, Crime, and Law, 14,* 19-28.

Davies, G.M., Stevenson-Robb, Y., & Flin, R. (1988). Tales out of school: Children's memory for an unexpected event. In M.M. Gruneberg, & P.E. Morris (Eds.). *Practical aspects of memory: Current research and issues, Vol.1 :Memory in everyday life* (pp. 122-127). Oxford: John Wiley & Sons.

Davies, G.M., Tarrant, A., & Flin, R. (1989). Close encounters of the witness kind: Children's memory for a stimulated health inspection. *British Journal of Psychology, 80,* 415-429.

Davies, G.M., Westcott, H.L., & Horan, N. (2000). The impact of questioning style on the content of investigative interviews with suspected child sexual abuse victims. *Psychology, Crime & Law, 6,* 81-97.

Davies, G.M., Wilson, C., Mitchell, R., & Milsom, J. (1995). *Videotaping children's evidence. An evaluation.* London: Home Office.

De Frias, C., Nilsson, L.-G., & Herlitz, A. (2006). Sex differences in cognition are stable over a 10-year period in adulthood and old age. *Aging, Neuropsychology, and Cognition, 13,* 574-587.

*Dekle, D.J., Beal, C.R., Elliott, R., & Huneycutt, D. (1996). Children as witnesses: A comparison of lineup versus showup identification methods. *Applied Cognitive Psychology, 10,* 1-12.

Dent, H. & Stephenson, G. (1979). An experimental study of the effectiveness of different techniques of questioning child witness. *British Journal of Social and Clinical Psychology, 18,* 41-51.

Diamond, R. & Carey, S. (1977). Developmental changes in the representation of faces. *Journal of Experimental Child Psychology, 23,* 1-22.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107-117.

Dingwall, R.G. (1980). Ethics and ethnography. *Sociological Review, 28*, 871-891.

Dodson, C.S., Johnson, M.K., & Schooler, J.W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition, 25,* 129-139.

Downing, P.E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293,* 2470-2473.

Doyle, J.M. (2001). Discounting the error costs: Cross-racial false alarms in the culture of contemporary criminal justice. *Psychology, Public Policy, and Law, 7,* 253-262.

Ellis, H.D. (1990). Developmental trends in face recognition. *The Psychologist, 3,* 114-119.

Ellis, H.D., Deregowski, J.B. & Shepherd, J.W. (1975). Descriptions of White and Black faces by White and Black subjects. *International Journal of Psychology, 10,* 119-123.

Ellis, H., Shepherd, J., & Bruce, A. (1973). The effects of age and sex upon adolescents' recognition of faces. *Journal of Genetic Psychology, 123,* 173-174.

Fallshore, M. & Schooler, J.W. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1608-1623.

Farah, M.J., Wilson, K.D., Drain, M. & Tanaka, J.N. (1998). What is "special" about face perception? *Psychological Review, 105,* 482-498.

Feinman, S. & Entwisle, D.R. (1976). Children's ability to recognize other children's faces. *Child Development, 47,* 506-510.

Fisher, G.H. & Cox, R.L. (1975). Recognizing human faces. *Applied Ergonomics, 6,* 104-109.

Fisher, R.P. & Geiselman. R.E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview.* Springfield, Illinois: Charles C. Thomas.

Flin, R., Boon, J., Bull, R., & Knox, A. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology, 83,* 323-337.

Flin, R.H. & Shepherd, J.W. (1986). Tell stories: Eyewitnesses' ability to estimate height and weight characteristics. *Human Learning, 5,* 29-38.

*Freire, A., Lee, K., Williamson, K.S., Stuart, S.J.E., & Lindasay, R.C.L. (2004). Lineup identification by children: Effects of clothing bias. *Law and Human Behavior, 28,* 339 – 354.

*Fritzley, V.H. & Lindsay, R.C.L. (2005). *Detecting innocence with a slow elimination lineup: An alternative for child witnesses.* Unpublished manuscript.

Galper, R.E. (1973). "Functional race membership" and recognition of faces. *Perceptual & Motor Skills, 37,* 455-462.

Gibson, E.J. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9,* 558-565.

Goetze, H.J. (1980). *The effect of age and method of interview on the accuracy and completeness of eyewitness accounts.* Unpublished doctoral dissertation, Hofstra University, Hempstead, NY.

Goldstein, A.G. (1977). The fallibility of the eyewitness: Psychological Evidence. In B.D. Sales (Ed.). *Psychology in the legal process* (pp. 223-247). New York: Spectrum.

Goldstein, A.G. (1979). Race-related variation of facial features: Anthropometric data I. *Bulletin of the Psychonomic Society, 13,* 187-190.

Goldstein, A.G. & Chance, J.E. (1971). Visual recognition memory for complex configurations. *Perception & Psychophysics, 9,* 237-241.

Goldstein, A.G. & Chance, J.E. (1979). Do "foreign" faces really look alike? *Bulletin of the Psychonomic Society, 13,* 111-113.

Goldstein, A.G. & Chance, J.E. (1980). Memory for faces and schema theory. *Journal of Psychology, 105,* 47-59.

Goodman, G. S. (1991). Commentary: On stress and accuracy in research on children's testimony. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 77-83). APA: Washington, DC.

Goodman, G. S., Aman, C., & Hirschman, J. (1987). Child sexual andphysical abuse: Children's testimony. In S. J. Ceci, M. P. Toglia, & D. F. Ross (Eds.), *Children's Eyewitness Memory* (pp. 1-24), New York: Springer-Verlag.

Goodman, G.S. & Aman, C. (1990). Children's use of anatomically detailed dolls to recount an event. *Child Development, 61,* 1859-1871.

*Goodman, G. S., Bottoms, B. L., & Schwartz-Kenney, B. M. (1991). Children's

testimony about a stressful event: Improving children's reports. *Journal of Narrative & Life History, 1,* 69-99.

*Goodman, G.S., Bottoms, B.L., Rudy, L., Davies, S.L., & Schwartz-Kenney, B.M. (2001). Effects of past abuse experiences on children's eyewitness memory. *Law and Human Behavior, 25,* 269-298.

*Goodman, G.S., Hirschman, J.E., & Hepps, D. (1991). Children's memory for stressful events. *Merrill-Palmer Quarterly, 37,* 109-157.

*Goodman, G.S. & Reed, R.S. (1986). Age differences in eyewitness testimony. *Law and Human Behavior, 10,* 317-332.

Goodman, G.S., Sayfan, L., Lee, J.S., Sandhei, M., Walle-Olsen, A., Magnussen, S., Pezdek, K., & Arredondo, P. (2007). The development of memory for own- and other-race faces. *Journal of Experimental Child Psychology, 98,* 233-242.

Greenhoot, A.F., Ornstein, P.A., Gordon, B.N., & Baker-Ward, L. (1999). Acting out the details of a pediatric check-up: The impact of interview condition and behavioral style on children's memory reports. *Child Development, 70,* 363-380.

Greenstock, J. & Pipe, M.-E. (1996). Interviewing children about past events: The influence of peer support and misleading questions. *Child Abuse & Neglect, 20,* 69-80.

*Gross, J. & Hayne, H. (1996). Eyewitness identification by 5- to 6-year-old children. *Law and Human Behavior, 20,* 359-373.

Gross, J. & Hayne, H. (1998). Drawing facilitates children's verbal reports of emotionally laden events. *Journal of Experimental Psychology: Applied, 4,* 163-179.

Gross, J. & Hayne, H. (1999a). Young children's recognition and description of their own and others' drawings. *Developmental Science, 2,* 476-489.

Gross, J. & Hayne, H. (1999b). Drawing facilitates children's verbal reports after long delays. *Journal of Experimental Psychology: Applied, 5,* 265-283.

Hay, D.C. & Cox, R. (2000). Development changes in the recognition of faces and facial features. *Infant and Child Development, 9,* 199-212.

Hershkowitz, I.U., Lamb, M.E., Sternberg, K.J., & Esplin, P.W. (1997). The relationships among interviewer utterance type, CBCA scores and the richness of children's responses. *Legal and Criminological Psychology, 2,* 169-176.

Holliday, R.E. & Albon, A.J. (2004). Minimising misinformation effects in young children with cognitive interview mnemonics. *Applied Cognitive Psychology, 18,* 263-281.

Home Office in conjunction with the Department of Health (1992). *Memorandum of good practice on video recorded interviews with children witnesses for criminal proceedings.* London: Her Majesty's Stationary Office.

Home Office (2002). *Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children.* London: Home Office Communication Directorate.

Horn, R. (1996). Negotiating research access to organizations. *The Psychologist, 9,* 551- 554.

Horton, C. & Smith, D.J. (1988). *Evaluating police work: An action research project.* London: Policy Studies Institute.

Hunt, J. (1984). The development of rapport though the negotiation of gender in field work among police. *Human Organisation, 43,* 283-296.

Hutcheson, G.D., Baxter, J.S., Telfer, K., & Warden, D. (1995). Child witness statement quality: Question type and errors of omission. *Law and Human Behavior, 19,* 631-648.

Hyde, J.S. & Linn, M.S. (1988). Gender differences in verbal ability. *Psychological Bulletin, 104,* 53-69.

Hyde, J.S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581-592.

Janssen, J.P. & Horowski, A. C. (1980). Students' estimates of stature: Tendency to accentuate as a cognitive style in person perception. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 12,* 167-176.

Jensen, A.R. (1998). *The g-factor: The science of mental ability (Human evolution, behavior, and intelligence).* Westport: Praeger Publishers.

Johnson, M.K., & Hashtroudi, S., & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3-28.

Jones, C.H. & Pipe, M.-E. (2002). How quickly do children forget events? A systematic study of children's event reports as a function of delay. *Applied Cognitive Psychology, 16,* 755 – 768.

Kask, K., Bull, R., Heinla, I., & Davies, G. (2007). The effect of a standard to improve person descriptions by children. *Journal of Police and Criminological Psychology, 22,* 77-83.

Kebbell, M., Milne, R., & Wagstaff, G. (1999). The cognitive interview: A survey of its forensic effectiveness. *Psychology, Crime, & Law, 5,* 101-115.

King, M.A. & Yuille, J.C. (1986). *An investigation of the witness abilities of children.* Unpublished manuscript, University of British Columbia.

Koppen, van P.J. & Lochun, S.K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior, 21,* 661- 685.

Korkman, J., Santtila, P., Drzewiecki, T., & Sandnabba, N.K. (2008). Failing to keep it simple: Language use in child sexual abuse interviews with 3-8-year-old children. *Psychology, Crime, and Law, 14,* 41-60.

Korkman, J., Santtila, P., & Sandnabba, N.K. (2006). Dynamics of verbal interaction between interviewer and child in interviews with alleged victims of sexual abuse. *Scandinavian Journal of Psychology, 47,* 109-119.

Korkman, J., Santtila, P., Westeraker, M. & Sandnabba, N.K. (2008). Interviewing techniques and follow-up questions in child sexual abuse interviews. *European Journal of Developmental Psychology, 5,* 108-128.

Kuehn, L.L. (1974). Looking down a gun barrel: Person perception and violent crime. *Perceptual and Motor Skills, 39,* 1159-1164.

La Rooy, D., Pipe, M.-E., & Murray, J.E. (2007). Enchancing children's event recall after long delays. *Applied Cognitive Psychology*, 21, 1-17.

Lamb, M.E. (1994). The investigation of child sexual abuse: An interdisciplinary consensus statement. *Child Abuse & Neglect, 18,* 1021-1028.

Lamb, M.E. & Fauchier, A. (2001). The effects of question type on self-contradictions by children in the course of forensic interviews. *Applied Cognitive Psychology, 15,* 483-491.

Lamb, M.E., Hershkowitz, I., Sternberg, K.J., Boat, B., & Everson, M.D. (1996). Investigative interviews of alleged sexual abuse victims with and without anatomic dolls. *Child Abuse and Neglect, 20,* 1239-1247.

Lamb, M.E., Hershkowitz, I.U., Sternberg, K.J., Esplin, P.W., Hovav, M., Manor,

T., & Yudilevitch, L. (1997). Effects of investigative utterance types on Israeli children's responses. *International Journal of Behavioral Development, 19,* 627-637.

Lamb, M.E., Orbach, Y., Sternberg, K.J., Esplin, P.W., & Hershkowitz, I. (2002). The effects of forensic interview practices on the quality of information provided by alleged victims of child abuse. In H. Westcott, G. Davies and R. Bull (Eds.) *Children's testimony: A handbook of psychological research and forensic practice* (pp. 131-146). Chichester: Wiley.

Lamb, M.E., Sternberg, K.J., & Esplin, P.W. (1998). Conducting investigative interviews of the alleged sexual abuse victims. *Child Abuse and Neglect, 22,* 813-823.

Lamb, M.E., Sternberg, K.J., & Esplin, P.W. (2000). Effects of age and delay on the amount of information provided by alleged sex abuse victims in investigative interviews. *Child Development, 71,* 1586-1596.

Lamb, M.E., Sternberg, K.J., Orbach, Y., Esplin, P.W., Stewart, H., & Mitchell, S. (2003). Age differences in young children's responses to open-ended invitations in the course of forensic interviews. *Journal of Consulting and Clinical Psychology, 71,* 926-934.

Lamb, M.E, Sternberg, K.J., Orbach, Y., Hershkowitz, I., & Esplin, P.W. (1999). Forensic interview of children. In A.Memon and R. Bull (Eds.). *Handbook of the psychology of interviewing* (pp. 253-277). New York: Wiley.

Lamb, M.E., Sternberg, K.J., Orbach, Y., Hershkowitz, I., Horowitz, D., & Esplin, P.W. (2002). The effects of intensive training and ongoing supervision on the quality of investigative interviews with alleged sex abuse victims. *Applied Developmental Science, 6,* 114-125.

Lee, R.M. (1993). *Doing research on sensitive topics.* London: Sage.

*Leippe, M.R., Romanczyk, A., & Manion, A.P. (1991). Eyewitness memory for a touching experience: Accuracy differences between child and adult witnesses. *Journal of Applied Psychology, 76,* 367-379.

Leventhal, J.M., Hamilton, J., Rededal, S., Tebano-Micci, A., & Eyster, C. (1989). Anatomically correct dolls used in interviews of young children suspected of having been sexually abused. *Pediatrics, 84,* 900-906.

Levi, A.M. (1998). Are defendants guilty if they were chosen in a lineup? *Law and Human Behavior, 12,* 389-407.

Levi, A.M. (2006). An analysis of multiple choices in MSL lineups, and a comparison with simultaneous and sequential ones. *Psychology, Crime & Law, 12,* 273-285.

Levin, D.T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 1364-1382.

Lewin, C. & Herlitz, A. (2002). Sex differences in face recognition: Women's faces make the difference. *Brain and Cognition, 50,* 121-128.

Lindholm, T. (2005). Own-age biases in verbal person memory. *Memory, 13,* 21-30.

Lindsay, D.S., Jack, P.C. & Christian, M.A. (1991). Other-race face perception. *Journal of Applied Psychology, 76,* 587-589.

*Lindsay, R.C.L. & Freire, A. (2001). Elimination lineups work poorly with younger children (under 8 years). Society for Applied Research in Memory and Cognition (SARMAC IV), Kingston, Ontario, June, 2001.

Lindsay, R.C.L. & Harvie, V. (1998). Hits, false alarms, correct and mistaken

identifications: The effects of method of data collection on facial memory. In M.Gruneberg, P.Morris, & R.Sykes (Eds.), *Practical aspects of memory: Currect research and issues, Vol 1, Memory in everyday life* (pp. 47-52). Chicester, England: Wiley

Lindsay, R.C.L., Lea, J.A., & Fulford, J.A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76,* 741-745.

Lindsay, R.C.L., Lea, J.A., Nosworthy, G.J., Fulford, J.A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased Lineups: Sequential Presentation Reduces the Problem. *Journal of Applied Psychology, 76,* 796-802.

Lindsay, R.C.L., Martin, R. & Webber, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law and Human Behavior, 18,* 527-541.

Lindsay, R.C.L., Nosworthy, G.J., Martin, R., Martynuck, C. (1994). Using mug shots to find suspects. *Journal of Applied Psychology, 79,* 121-130.

*Lindsay, R.C.L., Pozzulo, J.D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21,* 391-404.

Lindsay, R.C. & Wells, G.L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70,* 556-564.

Lipsey, M. & Wilson, D. (2001). *Practical meta-analysis.* Sage Publications.

Lyle, K.B. & Johnson, M.K. (2004). Effects of verbalization on lineup face recognition in an interpolated inspection paradigm. *Applied Cognitive Psychology, 18,* 393 – 403.

Malpass, R.S. (1981). Training in face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 271–285). London: Academic Press.

Malpass, R.S. (1990). An excursion into utilitarian analyses, with side trips. *Behavioral Science Research, 24,* 1-15.

Malpass, R.S. (1996). Enhancing eyewitness memory. In S.L.Sporer, R.S. Malpass, & G. Koehnken (Eds.). *Psychological issues in eyewitness identification* (pp. 177-204). Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Manis, M. & Paskewitz, J. (1984). Judging psychopathology: Expectation and contrast. *Journal of Personality and Social Psychology, 20,* 363-381. .

*Marin, B.V., Holmes, D.L., Guth, M., & Kovac, P. (1979). The potential of children as eyewitnesses. *Law and Human Behavior, 3,* 295-305.

Mauldin, M.A. & Laughery, K.R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology, 66,* 351-357.

McKelvie, S.J. (1981). Sex differences in memory for faces. *Journal of Psychology: Interdisciplinary and Applied, 107,* 109-125.

McQuiston-Surrett, D., Malpass, R.S., & Tredoux, C.G. (2006). Sequential vs. simultaneous lineups: A review of methods, data, and theory. *Psychology, Public Policy, and Law, 12,* 137-169.

Meissner, C.A. (2002). Applied aspects of the instructional bias effect in verbal overshadowing. *Applied Cognitive Psychology, 16,* 911-928.

Meissner, C.A. & Brigham, J.C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15,* 603 – 616.

Meissner, C.A., Brigham, J.C., & Kelley, C.M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition, 29,* 176-186.

Meissner, C.A., Sporer, S.L., & Schooler, J.F. (2006). Person descriptions as eyewitness evidence. In R.C.L.Lindsay, D.F.Ross, J.D.Read, & M.P.Toglia (Eds.), *Handbook of eyewitness psychology: Memory for People* (pp. 3-34). Mahwah, NJ: Lawrence Erlbaum.

Melnyk, L. & Bruck, M. (2004). Timing moderates the effects of repeated suggestive interviewing on children's eyewitness memory. *Applied Cognitive Psychology, 18,* 613 – 631.

Memon, A. & Bartlett, J.C. (2002) The effects of verbalisation on face recognition. *Applied Cognitive Psychology, 16,* 635-650.

Memon, A., Holley, A., Wark, L., Bull, R. & Köhnken, G. (1996). Reducing suggestibility in child witness interviews. *Applied Cognitive Psychology, 10,* 503-518.

*Memon, A. & Rose, R. (2002). Identification abilities of children: Does a verbal description hurt face recognition? *Psychology, Crime & Law, 8,* 229-242.

Memon, A. & Vartoukian, R. (1996). The effects of repeated questioning on young children's eyewitness testimony. *British Journal of Psychology, 87,* 403-416.

Milne, R. & Bull, R. (1999). *Investigative interviewing: Psychology and practice.* Chichester: John Wiley & Sons.

Milne, R. & Bull, R. (2003). Does the cognitive interview help children to resist the effects of suggestive questioning? *Legal and Criminological Psychology, 8,* 21-38.

Mosteller, F. & G.A. Colditz, G.A. (1996) Understanding research synthesis (meta-

analysis). *Annual Review of Public Health, 17,* 1–23.

Mundlak, Y., & Yahav, J. A. (1981). Random effects, fixed effects, convolution, and separation. *Econometrica,* 49, 1399-1416.

Newel, F.N., Chiroro, P., & Valentine, T. (1999). Recognizing unfamiliar faces: The effects of distinctiveness and view. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 52A,* 509-534.

Ng, W. & Lindsay, R.C.L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology, 25,* 217-232.

Nikonova, O. & Ogloff, J.R.P. (2005). Mock jurors' perceptions of child witnesses: The impact of judicial warning. *Canadian Journal of Behavioural Science, 37,* 1-19.

Ochsner, J.E., & Zaragoza, M.S. (1988). The accuracy and suggestibility of children's memory for neutral and criminal eyewitness events. Paper presented at the American Psychology and Law Meetings, Miami, FL.

Odinot, G. & Wolters, G. (2006). Repeated recall, retention interval and the accuracy-confidence relation in eyewitness memory. *Applied Cognitive Psychology, 20,* 973-985.

Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged abuse victims. *Child Abuse and Neglect, 24,* 733-752.

Orbach, Y. & Lamb, M.E. (1999). Assessing the accuracy of a child's account of sexual abuse: A case study. *Child Abuse and Neglect, 23,* 91-98.

Ornstein, P.A., Gordon, B.N., & Larus, D.M. (1992). Children's memory for a

personally experienced event: Implications for testimony. *Applied Cognitive Psychology, 6,* 49-60.

Paivio, A (1969). Mental imagery in associative learning and memory. *Psychological Review, 76,* 241-263.

Paivio, A. (1971). *Imagery and Verbal Processes*. New York: Holt, Rinehart & Winston.

*Parker, J.F. & Carranza, L.E. (1989). Eyewitness testimony of children in target-present and target-absent lineups. *Law and Human Behavior, 13,* 133-149.

*Parker, J.F., Haverfield, E., & Baker-Thomas, S. (1986). Eyewitness testimony of children. *Journal of Applied Social Psychology, 16,* 287-302.

*Parker, J.F. & Myers, A. (2001). Attempts to improve children's identifications from sequential-presentation lineups. *Journal of Applied Social Psychology, 31,* 796-815.

*Parker, J.F. & Ryan, V. (1993). An attempt to reduce guessing behavior in children's and adults' eyewitness identifications. *Law and Human Behavior, 17,* 11-26.

Pellicano, E. & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological Science, 14,* 618-622.

Peters, D.P. (1987). The impact of naturally occurring stress on children's memory. In S. J. Ceci, M. P. Toglia, & D. F. Ross (Eds.), *Children's eyewitness memory* (pp. 122-141). New York: Springer-Verlag

Peters, D.P. (1991). The influence of stress and aruousal on the child witness. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 60-77). APA: Washington, DC.

Peterson, C. (1996). The preschool child witness: Errors in accounts of traumatic injury. *Canadian Journal of Behavioural Science, 28,* 36-42.

*Pezdek, K., Blandon-Gitlin, I., & Moore, C. (2003). Children's face recognition memory: More evidence for the cross-race effect. *Journal of Applied Psychology, 88,* 760-763.

Pickel, K.L., Narter, D.B., Jameson, M.M., & Lenhardt, T.T. (2007). The weapon focus effect in child eyewitnesses. *Psychology, Crime, and Law, 14,* 61-72.

Pigott, M. & Brigham, J.C. (1985). Relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology, 70,* 547-555.

Pipe, M.-E., Sutherland, R., Webster, N., Jones, C., & La Rooy, D. (2004). Do early interviews affect children's long-term event recall? *Applied Cognitive Psychology, 18,* 823 – 839.

Platz, S.J. & Hosch, H.M. (1988). Cross-racial/ethnic eyewitness identification: A field study. *Journal of Applied Social Psychology, 18,* 972-984.

Poole, D.A. & Lamb, M.E. (1998). *Investigative interviews of children: A guide for helping professionals.* Washington, DC: APA.

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,* 844-853.

Powers, P.A., Andriks, J.L. & Loftus, E.F. (1979). The eyewitness accounts of females and males. *Journal of Applied Psychology, 64,* 339-347

*Pozzulo, J.D. & Dempsey, J. (2006). Biased lineup instructions: Examining the effect of pressure on children's and adults' eyewitness identification accuracy. *Journal of Applied Social Psychology, 36,* 1381-1394.

Pozzulo, J.D., Lemieux, J.M.T., Wells, E., & McCuaig, H.J. (2007). The influence of eyewitness identification decisions and age of witness on jurors' verdicts and perceptions of reliability. *Psychology, Crime, & Law, 12,* 641-652.

*Pozzulo, J.D. & Lindsay, R. C. L. (1997). Increasing correct identifications by children. *Expert Evidence, 5,* 126-132.

Pozzulo, J.D. & Lindsay, R. C. L. (1998). Identification accuracy of children versus adults: A meta-analysis. *Law and Human Behavior*, *22,* 549-570.

*Pozzulo, J.D. & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. *Journal of Applied Psychology, 84,* 167-176.

*Pozzulo, J.D. & Warren, K.L. (2003). Descriptions and identifications of strangers by youth and adult eyewitnesses. *Journal* of *Applied Psychology, 88,* 315-323.

Price, D.W. & Goodman, G.S. (1990). Visiting the wizard: Children's memory for a recurring event. *Child Development, 61,* 664-680.

Rachlew, A. (2007). Photo, video or live identification parades: Which procedure provides the most reliable result when police officers conduct the interviews? Paper presented in 4th conference of Nordic Network of Research on Psychology and Law (NNPL). Reykjavik, October.

Rehnman, J. (2007). *The role of gender in face recognition.* Doctoral dissertation: Stockholm University.

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology*, *80,* 313-331.

*Ricci, C.M. & Beal, C.R. (1998). Effect of questioning techniques and interview setting on young children's eyewitness memory. *Expert Evidence, 6,* 127-144.

*Ricci, C.M., Beal, C.R., & Dekle, D.J. (1996). The effect of parent versus unfamiliar interviewers on children's eyewitness memory and identification accuracy. *Law and Human Behavior, 20,* 483-500.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage

Ross, D. F., Dunning, D., Toglia, M. P. & Ceci, S. J. (1990). The child in the eyes of the jury: Assessing the mock jurors' perceptions of the child witness. *Law and Human Behavior, 14,* 5-23.

Sangrigoli, S. & de Schonen, S. (2004). Recognition of own-race and other-race faces by three-month-old infants. *Journal of Child Psychology and Psychiatry, 45,* 1219-1227.

Sangrigoli, S., Pallier, C., Argenti, A.-M., Ventureyra, V.A.G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science, 16,* 440-444.

Santtila, P., Korpela, S., & Häkkänen, H. (2004). Expertise and decision-making in the linking of car crime series. *Psychology, Crime and Law, 10,* 97-112.

Saywitz, K.J. & Goodman, G.S. (1996). Interviewing children in and out of court: Current research and practice implications. In J. Briere, L. Berliner, J. A. Bulkley, C. Jenny, & T. Reid (Eds.), *The APSAC handbook on child maltreatment* (pp. 297-318). Thousand Oaks, CA: Sage.

Saywitz, K. J., Goodman, G. S., Nicholas, E., & Moan, S. F. (1991). Children's memories of a physical examination involving genital touch: Implications for reports of child sexual abuse. *Journal of Consulting and Clinical Psychology, 59,* 682-691.

Schepard, A.I. (2004). *Children, courts, and custody: Interdisciplinary models for divorcing families.* New York, US: Cambridge University Press.

Schooler, J.W. & Engster-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22,* 36-71.

Schretlen, D.J., Pearlson, G.D., Anthony, J.C., & Yates, K.O. (2005). Determinants of Benton facial recognition test: Performance in normal adults. *Neuropsychology, 15,* 405-410.

*Schwartz-Kenney, B. M., Bottoms, B. L., & Goodman, G. S. (1996). Improving children's person identification. *Child Maltreatment: Journal of the American Professional Society on the Abuse of Children, 1,* 121-133.

Schwarzer, G. & Massaro, D.W. (2001). Modeling face identification processing in children and adults. *Journal of Experimental Child Psychology, 79,* 139-161.

Searcy, J., Bartlett, J.C., & Memon, A. (2000). Influence of post-event narratives, line-up conditions and individual differences on false identification by young and older eyewitnesses. *Legal and Criminological Psychology, 5,* 219-235.

Seitz, K. (2002). Parts and wholes in person recognition: Developmental trends. *Journal of Experimental Child Psychology, 82,* 367-381.

Shapiro, P.N. & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139-156.

Shepherd, J.W. (1981). Social factors in face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 55–79). London: Academic Press.

Shepherd, J.W. (1983). Identification after long delays. In S. M. A. Lloyd-Bostock & R. B. Clifford (Eds.), *Evaluating witness evidence* (pp. 173-197). London: John Wiley & Sons.

Shepherd, J.W. & Deregowski, J.B. (1981). Races and faces - A comparison of the responses of Africans and Europeans to faces of the same and different races. *British Journal of Social Psychology, 20,* 125-133.

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. 87-116). Mahwah, New Yersey: Lawrence Erlbaum Associates.

Slater, A., Bremner, G., Johnson, S. P., Sherwood, P., Hayes, R., & Brown, E. (2000). Newborn infants' preference for attractive faces: The role of internal and external facial features. *Infancy, 1,* 265-274.

Slone, A.E., Brigham, J.C. & Meissner, C.A. (2000). Social and cognitive factors affecting the own-race bias in Whites. *Basic & Applied Social Psychology, 22,* 71-84.

Sporer, S.L. (1989). Verbal and visual processes in person identification. In H. Wegener, F. Loesel, & J. Haisch (Eds.), *Criminal behavior and the justice system: Psychological perspectives* (pp. 303-324). New York: Springer.

Sporer, S.L. (1992a). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22,* 157-180.

Sporer, S.L. (1992b, March). An archival analysis of person descriptions. Paper

presented at the Biennial Meeting of the American Psychology-Law Society in San Diego, California.

Sporer, S.L. (1996). Psychological aspects of person descriptions. In S.L.Sporer, R.S.Malpass, & G.Koehnken (Eds.), *Psychological Issues in Eyewitness Identification* (pp. 53-86). New Jersey: Lawrence Erlbaum Associates

Sporer, S.L. (2001a). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, & Law, 7,* 36-97.

Sporer, S.L. (2001b). The cross-race bias: Beyond recognition of faces in the laboratory. *Psychology, Public Policy, & Law, 7,* 170-200.

Sporer, S.L. (2007). Person descriptions as retrieval cues: Do they really help? Psychology, Crime, and Law, iFirst, 1-19. Retrieved 19.11.2007 from http://www.informaworld.com/smpp/title~content=g769588626~db=all.

Sporer, S.L., Trinkl, B., & Guberova, E. (2007). Matching faces: Differences in Processing speed of out-group faces by different ethnic groups. *Journal of Cross-Cultural Psychology, 38,* 398-412.

Statistics Canada. (31 October 2001). Homicide statistics. *The Daily.* Retrieved 22.08.2002 from http://www.statcan.ca/Daily/English/011031/d011031b.htm.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R.C.L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25,* 459-473.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R.C.L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27,* 523 – 540.

Sternberg, K.J., Lamb, M.E., Davies, G.M., & Westcott, H.L. (2001). The

Memorandum of Good Practice: Theory versus application. *Child Abuse & Neglect, 25,* 669-681.

Sternberg, K.J., Lamb, M.E., Hershkowitz, I., Esplin, P.W., Redlich, A., & Sunshine, N. (1996). The relation between investigative utterance types and the informativeness of child witnesses. *Journal of Applied Developmental Psychology, 17,* 439-451.

Sternberg, K.J., Lamb, M.E., Hershkowitz, I., Yudilevitch, L., Orbach, Y., Esplin, P.W., & Hovev, M. (1997). Effects of introductory style on children's abilities to describe experiences of sexual abuse. *Child Abuse & Neglect, 21,* 1133-1146.

Stockdale, M.S. (1996). *Sexual harassment in the workplace: Perspectives, frontiers, and response strategies.* Thousand Oaks, CA: Sage.

Strange, D., Garry, M., & Sutherland, R. (2002). Drawing out children's false memories. *Applied Cognitive Psychology, 17,* 607 – 619.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science: A Multidisciplinary Journal, 12,* 257-285.

Swope, T.M. (1994). *Social experience, illusory correlation and facial recognition ability.* Unpublished master's thesis, Florida State University.

Tanaka, J.W., Kay, J.B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition, 5,* 479-496.

Taylor, M.J., Edmonds, G.E., McCarthy, G., & Allison, T. (2001). Eyes first! Eye processing develops before face processing in children. *Cognitive Neuroscience & Neuropsychology, 12,* 1671-1676.

Thierry, K.L., Lamb, M.E., Orbach, Y., & Pipe, M.-E. (2005). Developmental

differences in the function and use of anatomical dolls during interviews with alleged sexual abuse victims. *Journal of Consulting and Clinical Psychology, 73,* 1125-1134.

Tobey, A.E. & Goodman, G.S. (1992). Children's eyewitness memory: Effects of participation and forensic context. *Child Abuse & Neglect, 16,* 779-796.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79,* 471-491.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race on face recognition. *Quarterly Journal of Experimental Psychology, 43A,* 161-204.

Valentine, T. (2007). The science of eyewitness identification evidence. Paper presented in the conference 'Forensic Psychology: Investigating, Identifying, Interviewing'. Leicester, June.

Valentine, T. & Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica, 61,* 259-273.

Valentine, T., Darling, S., & Memon, A. (2007). Do strict rules and moving images increase the reliability of sequential identification procedures? *Applied Cognitive Psychology, 21,* 933-949.

Valentine, T. & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology, 44A,* 671-703.

Walker, N.E. & Hunt, J.S. (1998). Interviewing child victim-witnesses: How you ask is what you get. In C. P. Thompson, D. J. Herrmann, J. D. Read, D. Bruce, D. G. Payne & M. P. Toglia (Eds.), *Eyewitness memory: Theoretical and applied perspectives* (pp. 55-87). Mahwah, NJ: Lawrence Erlbaum Associates.

Walker, P.M. & Hewstone, M. (2006). A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology, 20,* 461-475.

Warren, A.R. & Lane, P. (1995). Effects of timing and type of questioning on eyewitness accuracy and suggestibility. In M.S. Zaragoza, J.R. Graham, G.C.N. Hall, R. Hirschman, & Y.S. Ben-Porath, (Eds.). *Memory and testimony in the child witness*. (pp. 44-60). Thousand Oaks, CA: Sage.

Warren, C.A.B. (1988). *Gender issues in field research*. London: Sage.

*Weber, N., Brewer, N., & Wells, G.L. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12-second rule. *Journal* of *Experimental Psychology*: *Applied, 10,* 139-147.

Wegner, M.J. & Ingvalson, E.M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 872-892.

Wells, G.L. (1984). The psychology of lineup identification. Jo*urnal of Applied Social Psychology*, *14,* 89-103.

Wells, G. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? *Journal of Applied Psychology, 70,* 619-626.

Wells, G.L., Charman, S.D., & Olson, E.A. (2005). Building face composites can harm lineup identification performance. *Journal of Experimental Psychology: Applied*, *11,* 147-156.

Wells, G.L., & Luus, C.A.E. (1990). The diagnosticity of a lineup should not be confused with the diagnostic value of non-lineup evidence. *Journal of Applied Psychology, 75,* 511-516.

Wells, G.L. & Olson, E.A. (2001). The other-race identification in eyewitness

identification: What do we do about it? *Psychology, Public Policy, and Law, 7(1),* 230-246

Wells, E.C. & Pozzulo, J.D. (2006). Accuracy of eyewitnesses with a two-culprit crime: Testing a new identification procedure. *Psychology, Crime & Law, 12,* 417-427.

Westcott, H., Davies, G. M. & Clifford, B. R. (1989). The use of anatomical dolls in child witness interviews: A review. *Adoption and Fostering, 13,* 6-14.

Westcott, H.L. & Kynan, S. (2004). The application of a 'story-telling' framework to investigative interviews for suspected child sexual abuse. *Legal and Criminological Psychology, 9,* 37-56.

Westcott, H.L. & Kynan, S. (2006). Interviewer practice in investigative interviews for suspected child sexual abuse. *Psychology, Crime & Law, 12,* 367-382.

Westenholz, A. (1993). Paradoxical thinking and change in the frames of reference. *Organization Studies, 14,* 37-58.

Wilcock, R.A., Bull, R., & Vrij, A. (2005). Aiding the performance of older eyewitnesses: Enhanced non-biased line-up instructions and line-up presentation. *Psychiatry, Psychology and Law, 12,* 129-140.

Wilcock, E., Morgan, K., & Hayne, H. (2006). Body maps do not facilitate children's reports of touch. *Applied Cognitive Psychology, 20,* 607-615.

Wogalter, M.S. (1991). Effects of post-exposure deception and imaging on subsequent face recognition performance. *Proceeding of the Human Factors Society, 35,* 575-579.

Wogalter, M.S. (1996). Describing faces from memory: Accuracy and effects of subsequent recognition performance. *Proceeding of the Human Factors Society, 40,* 536-540.

Wogalter, M.S., Malpass, R.S., & Berger, M.A. (1993). How police officers construct lineups: A national survey. In *Proceedings of the Human Factors and Ergonomics Society* (pp.640-644). The Human Factors and Ergonomics Society: Santa Monica, CA.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis.* Newbury Park, CA: Sage

Wood, J. M., Schreiber, N., Martinez, Y., McLaurin, K., Strok, R., Velarde, L. D., Garven, S., & Malpass, R. S. (1998, March). Child interviewing techniques in the McMartin Preschool and Kelly Michaels cases: A quantitative comparison. Paper presented at the biennial conference of the American Psychology–Law Society, Redondo Beach, CA.

Wright, D. B., Boyd, C. E. & Tredoux, C. G. (1999, June). Own-race bias in South Africa and England. Paper presented at the biennial meeting of the Society of Applied Research in Memory & Cognition, Boulder, CO.

Wright, D.B. & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica, 114,* 101-114.

Wright, D.B. & Stroud, J.N. (2002). Age differences in lineup identification accuracy: People are better with their own age. *Law and Human Behavior, 26,* 641 – 654.

Wright, R., Powell, M.B., & Ridge, D. (2007). What criteria do police officers use to measure the success of an interview with a child? *Psychology, Crime & Law, 13,* 395-404.

Yarmey, A.D. (1993). Adult age and gender differences in eyewitness recall in field settings. *Journal of Applied Social Psychology, 23,* 1921-1932.

Yarmey, A.D. & Jones, H.P (1983). Accuracy of memory of male and female

eyewitnesses to a criminal assault and rape. *Bulletin of Psychonomic Society, 21,* 89-92.

Yarmey, A.D. & Yarmey, M.J. (1997). Eyewitness recall and duration estimates in field settings. *Journal of Applied Social Psychology, 27,* 330-344.

Yarmey, A.D., Yarmey, A.L., & Yarmey, M.J. (1994). Face and voice identifications in showups and lineups. *Applied Cognitive Psychology, 8,* 453-464.

Yuille, J.C. & Cutshall, J.L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology, 71,* 291-301.

Zajac, R. & Hayne, H. (2003). I don't think that's what really happened: The effect of cross-examination on the accuracy of children reports. *Journal of Experimental Psychology: Applied, 9,* 187-195.

Zajac, R. & Hayne, H. (2006). The negative effect of cross-examination style questioning on children's accuracy: Older children and not immune. *Applied Cognitive Psychology, 20,* 3-16.

9. APPENDICES

9.1. Appendix 1: Questions asked using or not using the standard in Study 3

No standard mode	Using standard
Was my friend a man or a woman?	I am X. Was my friend a man or a woman?
Show me on this scale, how tall was my friend?	I am this tall (stands beside the scale). Show me on this scale, how tall was my friend?
How old was my friend?	I am X years old. Compared to me. how old was my friend?
Look, in here different people are drawn here. Show me, whom did my friend look similar to?	Look, different people are drawn here. I would be similar to this person here. Show me, with whom did my friend look similar to?
There are some faces on this paper. Which face did my friend have?	There are some faces on this paper. I have this kind of face. Which face did my friend have?
There are some hairstyles on this paper. Which hairstyle did my friend have?	The are some hairstyles on this paper. My hair looks like this. Which hairstyle did my friend have?
There are some mouths on this paper. Which mouth did my friend have?	There are some mouths on this paper. I have this kind of mouth. Which mouth did my friend have?
There are some noses on this paper.	There are some noses on this paper.

Which nose did my friend have?	I have this kind of nose. Which nose did my friend have?
Did he have big, small or average nose?	I have X nose. Did he have big, small or average nose?
Did he have straight or hooked nose?	I have X nose. Did he have straight or hooked nose?
Did he have big, small or average ears?	I have X ears. Did he have big, small or average ears?
Try to remember what colour were his eyes?	I have grey eyes. Try to recall what colour were his eyes.
What was his hair colour?	My hair is black. What was his hair colour?
Show how long his hair was?	My hair is this long. Show how long his hair was?
What clothes did he have on?	I wear this and this. What clothes did he have on?
What footwear did he have?	I wear this and this. What footwear did he have?
Do you remember anything else about him?	Do you remember anything else about him?

NB! For each mentioned piece of clothing – what was its' colour?

9.3. Appendix 3: Targets and their target replacements (below each) in Study 4

Targets (from left to right: Afro-American, Latino, Caucasian, Turkish, Chinese).

Target replacements (from left to right: Afro-American, Latino, Caucasian, Turkish, Chinese).

9.4. Appendix 4: Stimulus sheets in Study 4

Stimulus sheet 1

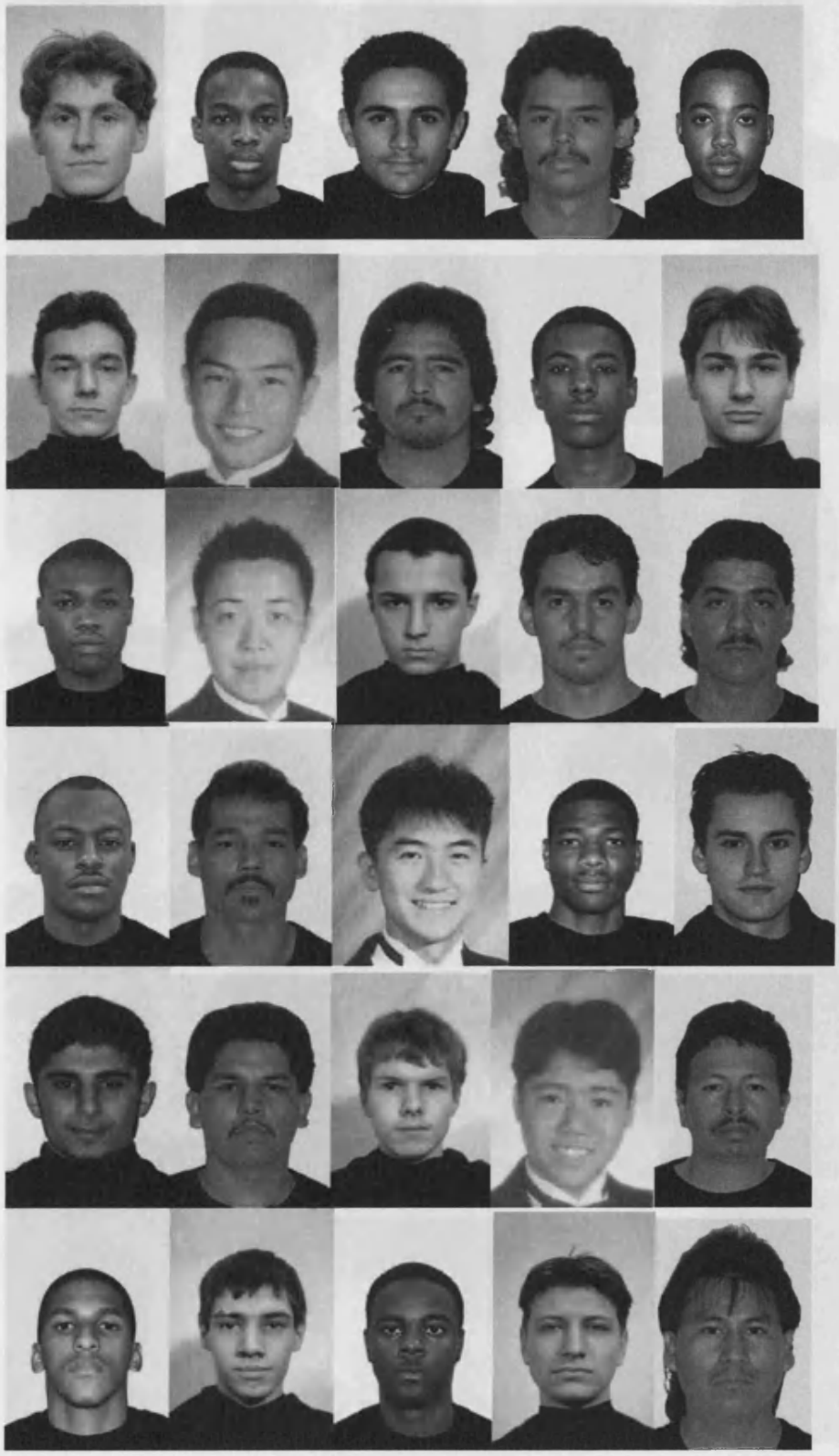
Stimulus sheet 2

Stimulus sheet 3

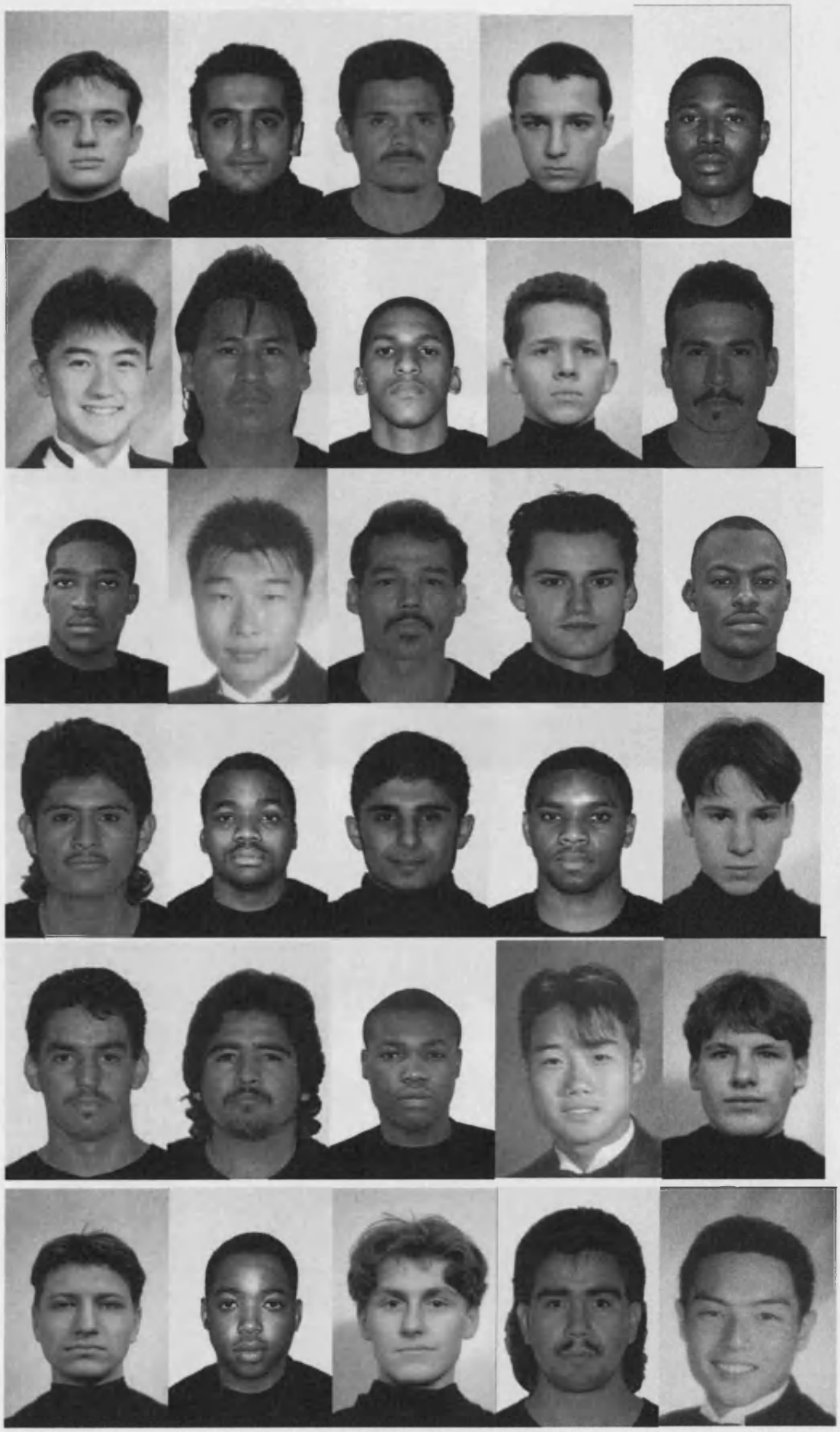
Stimulus sheet 4

9.5. Appendix 5: Face recognition sets in Study 4.

Face recognition set 1.



Face recognition set 2.



9.6. Appendix 6: Targets and their target replacements (below each) in Study 5

Targets

Target replacements

9.7. Appendix 7: Stimulus sheets in Study 5

Stimulus sheet 1

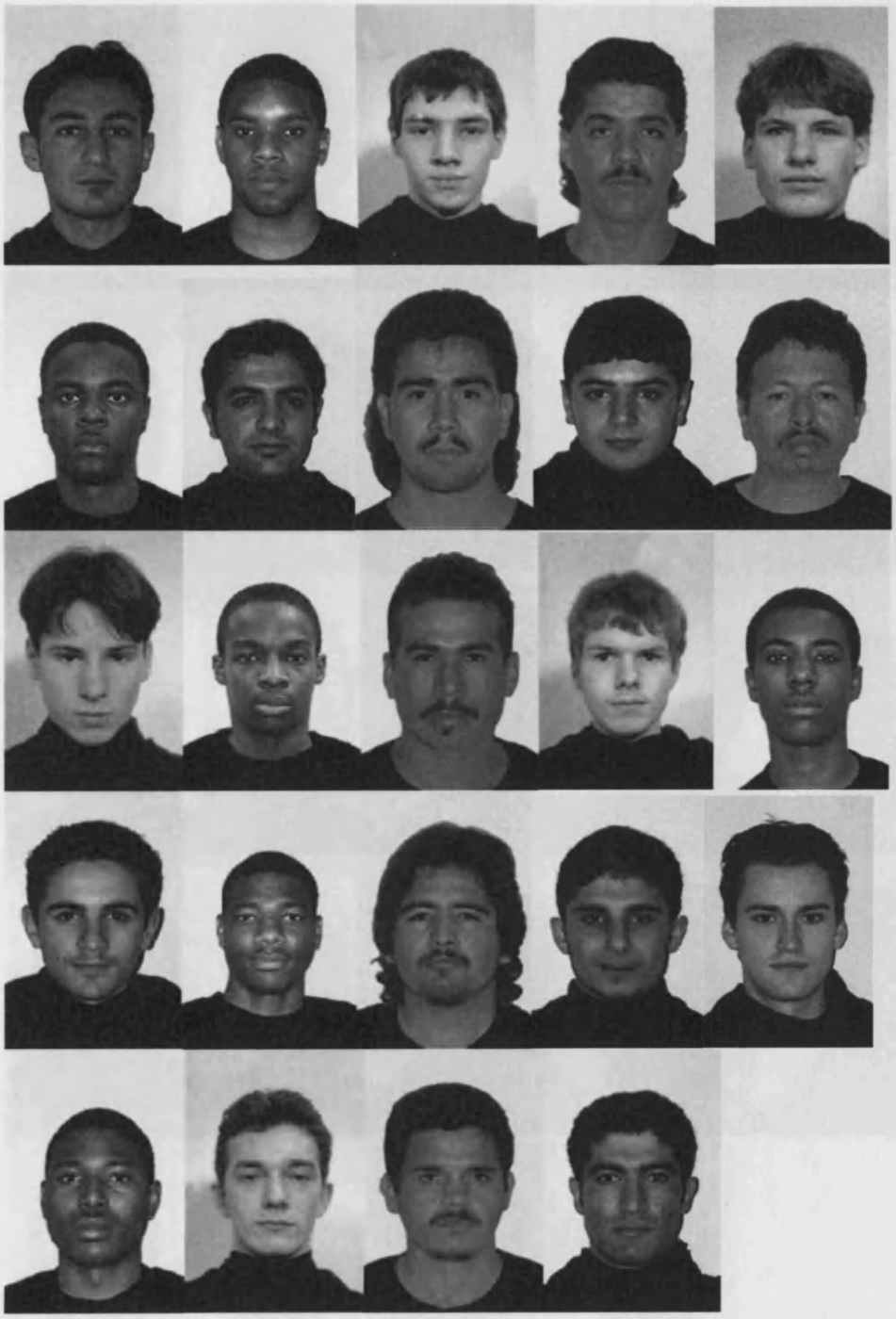
Stimulus sheet 2

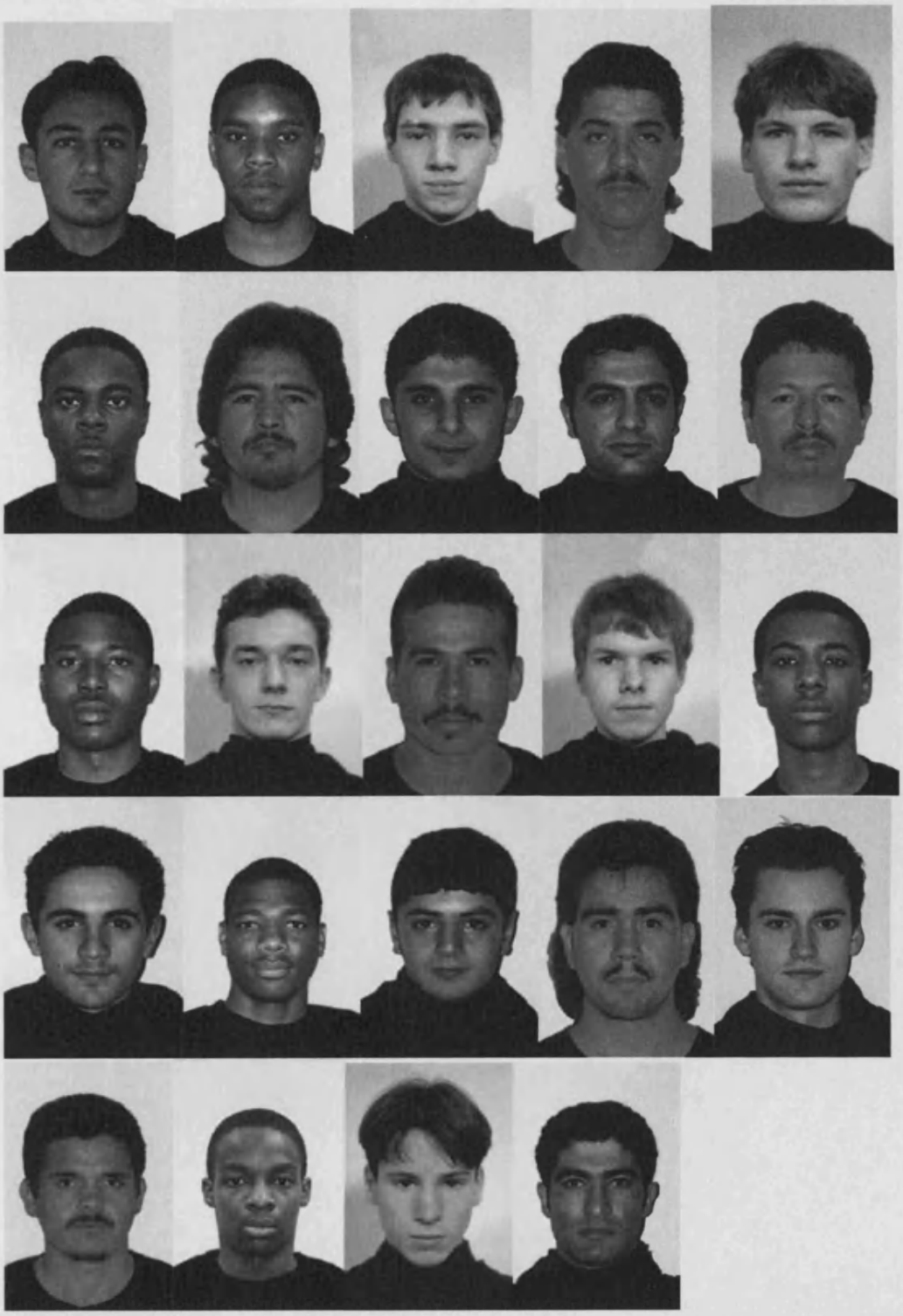
Stimulus sheet 3

Stimulus sheet 4

9.8. Appendix 8: Face recognition sets in Study 5.

Face recognition set 1 (used in ′decide all′ and ′decide some` presentation methods, presented sequentially one at the time from left to right)



Face recognition set 2 (used in ´decide all´ and ´decide some` presentation methods, presented sequentially one at the time from left to right).

Note. Sequential lineups were comprised of either target and five foils (in TP mode) or target replacement and five foils (in TA mode).