

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Comparing featural and holistic composite systems with the aid of Guided Memory techniques

Taryn Sullivan

Submitted in partial fulfillment of the requirements for the
award of the degree of Master of Arts in Psychological
Research

Faculty of Humanities
UNIVERSITY OF CAPE TOWN
February 2007

Declaration

I, Taryn Sullivan, declare that this dissertation is my own work. This work has not been previously submitted in whole, or in part, for the award of any degree. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

signature removed

Taryn Sullivan

Acknowledgements

"It takes both rain and sunshine to make a rainbow"

Unknown

There are many people who have been of great assistance to me during my studies. Firstly, very special thanks go to the special people in my life, who made this research process a pleasant one: to my parents for all their loving support and encouragement throughout my journey of obtaining a MA degree, and to all my close friends who were there for me during the stressful times. Most importantly to the amazing man in my life, Jacques, for motivating me every day to be better, try harder and for always keeping a smile on my face.

I would also like to thank my supervisor, Professor Colin Tredoux for his constant support and patience, invaluable insights and suggestions, even when thousands of kilometers away. I'd like to thank the ACSENT lab for use of the facilities when conducting the experiments and to the many lab members for their comments, suggestions, thoughts and ideas, especially: Dave Nunez, Susan Malcolm-Smith, Heike Schmidt, Nils Maskow, Katharina Zimmer and Peter Schaupp.

A particular thanks goes to Dave for setting up my evaluation tasks online and for all the help with the statistical analyses and overall advice. Thanks also to Heike, Nils and Katharina for their help in collecting the many photographs used for the ID database and the time-consuming landmarking of these – your help really sped up the research process and made my life a lot easier during very hectic times. Thanks must also be given to all those UCT students who volunteered their faces to be photographed and used in the database.

Last but certainly not least, I would like to gratefully thank the National Research Foundation (Grant-Holder linked bursary) and University of Cape Town (Siri Johnson and Jock Beattie Memorial Bursaries) for partial funding of this research.

Honouring the beautiful memory that was Jordan Leigh ...

Abstract

This study compares the effectiveness of two computerised composite construction systems – a holistic, recognition-based system named ID and a featural system that is utilized internationally, namely FACES. The comparison aimed to test whether ID produces better quality composites to FACES, and whether these composites could be improved with the aid of context reinstatement techniques, in particular guided memory.

Participants (n=64) attended a staged event where they witnessed a female 'numerologist' for 20 minutes. Five weeks later they were asked to return to create a composite of the woman using either FACES or ID. Reconstructions were made in view, from memory after a South African Police interview or from memory after a guided memory interview. In addition, experts for each system constructed composites of each perpetrator. Studies have reported enhanced identification when multiple composites are combined to create a morph. Hence, the guided memory composites for each perpetrator were morphed to create three ID and three FACES morphs. The complete set of 76 composites was then evaluated by 503 independent judges using matching and rating tasks.

The study hypothesised that ID would perform better, but results suggest that the two systems performed equivalently. Results also suggest that the guided memory interview did not have the desired effect of significantly improving participants' memories of the perpetrator, and that contrary to expectations, the morphed composites performed extremely poorly and were rated the worst and identified the least. Related findings and ideas for future research are discussed.

Contents

Declaration.....	1
Acknowledgements.....	2
Abstract.....	3
Contents.....	4
List of Figures.....	6
List of Tables.....	8
Chapter 1 Introduction.....	10
Chapter 2 Review of the Literature.....	15
1. Methods of facial composite production.....	15
Verbal descriptions.....	15
Sketch artistry.....	18
Manual and computerised systems.....	19
2. Why are face recall systems so limited?.....	30
3. Face processing theories.....	32
Featural face processing.....	32
Holistic face processing.....	34
Dual face processing.....	36
4. New generation configural composite systems: "Evolving faces".....	39
EvoFit.....	39
New computerised system: ID.....	43
The comparative system in this study: FACES 4.0.....	52
5. Context reinstatement.....	55
Guided Memory techniques.....	56
The Cognitive Interview.....	61
6. Multiple composites and morphing techniques.....	65
Chapter 3 Methodology.....	68
Stage 1: Database building.....	68
Stage 2: Event phase.....	74
Method.....	74

Procedure.....	75
Stage 3: Reconstruction and morphing phase.....	77
Method	77
Procedure.....	86
Morphing.....	89
Stage 4: Evaluation phase.....	91
Method	91
Procedure.....	94
Chapter 4 Results.....	95
1. Evaluation identifications.....	97
Effect of Guided Memory on identification rates.....	97
Differences between the composite systems.....	100
Interaction between the conditions and systems.....	101
Gender, race and task order effects on identification rates.....	104
Rates of choosing of lineup members.....	106
Conclusion	108
2. Evaluation ratings.....	109
Race and gender effects on similarity ratings.....	109
Interaction between the conditions and systems.....	113
Conclusion	118
3. Evaluation of the guided memory interview.....	119
4. Familiarity and lineup identifications resulting from the post- experimental questioning.....	120
Chapter 5 Discussion.....	122
1. Does ID produce better quality composites than FACES?.....	122
2. Does context reinstatement improve memory?.....	126
3. Do morphs produce a 'super composite'?.....	132
Chapter 6 General conclusion.....	134
References.....	138
Appendix A Photo protocol.....	148
Appendix B Perpetrator photographs and the 76 composites constructed.....	150
Appendix C South African Police interview used in Study 1.....	153
Appendix D Lineups used for the evaluation tasks.....	154

List of Figures

Figure 1: The main ID graphical user interface (GUI) during a reconstruction	43
Figure 2: The optimise feature(s) tool.....	46
Figure 3: Effects of the optimise features tool on the eyes.....	46
Figure 4: The main FACES 4.0 user interface during a reconstruction	52
Figure 5: Graphical representation of the camera setup protocol used.....	69
Figure 6: An example of incorrect (a) and correct (b) facial alignment	71
Figure 7: An example of hair that could not be edited out (face excluded from the database).....	71
Figure 8: An example of the face outline path being applied and altered to fit the face. The handles seen in (b) are used to control the curve of the path, allowing precision accuracy.....	72
Figure 9: A face before and after landmarking. Note the hair has been cropped to a tied back style.....	72
Figure 10: The main interface of the FaceWarp programme in use. The image to the left is the landmarked photograph (inputted image); the image to the right is the warped average of the inputted image.....	73
Figure 11: The first four slides of the numerology presentation.....	75
Figure 12: A 3-view lineup for perpetrator one (lineup 1)	79
Figure 13: Examples of FACES composites before (left) and after (right) editing the hair style in Adobe Photoshop	88
Figure 14: Three individual ID guided memory composites (top), the morph (bottom left) and the perpetrator (bottom right)	90
Figure 15: Two individual FACES guided memory composites (top), the morph (bottom left) and the perpetrator (bottom right)	91
Figure 16: Examples of the evaluation tasks. Top = similarity rating; bottom = lineup identification after a first choice has been selected	93
Figure 17: Interaction identification rates for perpetrator one	101
Figure 18: Interaction identification rates for perpetrator two.....	101
Figure 19: Interaction identification rates for perpetrator three.....	102

Figure 20: Percentage identification rates for each perpetrator across gender of reconstructor	103
Figure 21: Identification rates for the ID and FACES systems across perpetrators for two memory conditions	104
Figure 22: Rates of choosing collapsed across perpetrators for FACES and ID	108
Figure 23: Similarity ratings across the three perpetrators for both race groups. Note: W = white; NW = not white.....	111
Figure 24: Similarity ratings across the three perpetrators for males and females.....	113
Figure 25: Perpetrator one similarity ratings across conditions for the two systems	114
Figure 26: Graph of cell mean profile of similarity ratings for each condition and system for perpetrator one	115
Figure 27: Graph of cell mean profile of similarity ratings for each condition and system for perpetrator two.....	116
Figure 28: Perpetrator two similarity ratings across conditions for the two systems	117
Figure 29: Perpetrator three similarity ratings across conditions for the two systems	118
Figure 30: The best identified composite (bottom) of each perpetrator (top) with percentage of correct identifications underneath	126

List of Tables

Table 1: Percentage of composites correctly matched to the corresponding target in a 6-person target present lineup.....	51
Table 2: Lineup fairness indices for each of the six lineups used in the evaluation tasks	80
Table 3: Overall percentages of hits across all three perpetrators. Number of participants is shown in brackets.....	98
Table 4: Percentage identification rates (based on judges' first choice selections) for each perpetrator across composite conditions. Number of participants is shown in brackets.....	99
Table 5: Percentage identification rates for each perpetrator across composite systems based on judges' first choice selections. Number of participants is shown in brackets.....	100
Table 6: Overall percentages of hits across all three perpetrators. Number of participants is shown in brackets.....	100
Table 7: Pearson Chi-square test of contingency results.	105
Table 8: Expected and observed frequencies for each perpetrator. Note: Observed = Raw cumulated frequencies	107
Table 9: Rates of choosing collapsed across perpetrators for the five construction conditions. Obs = observed (raw cumulated) frequencies; Exp = expected frequencies.....	108
Table 10: Comparative similarity ratings of composites across perpetrators and race of judges (a = Mean; b = Std. Dev.; c = n).....	110
Table 11: ANOVA summary table	110
Table 12: Comparative similarity ratings of composites across perpetrators and gender of judges (a = Mean; b = Std. Dev.; c = n).....	112
Table 13: ANOVA summary table	112
Table 14: ANOVA summary table	114
Table 15: Comparative similarity ratings of composites across condition and system for perpetrator one (a = Mean; b = Std. Dev.; c = n)	114

Table 16: ANOVA summary table	115
Table 17: Comparative similarity ratings of composites across condition and system for perpetrator two (a = Mean; b = Std. Dev.; c = n).....	116
Table 18: ANOVA summary table	117
Table 19: Comparative similarity ratings of composites across condition and system for perpetrator three (a = Mean; b = Std. Dev.; c = n).....	117
Table 20: Means and standard deviations for construction conditions for Study 1. (gm=guided memory; iv=in-view; sap=police interview)	119
Table 21: ANOVA summary table	119

University of Cape Town

Chapter 1

Introduction

“The human ability to recognize faces is remarkable. We can recognize thousands of faces learned throughout our lifetime and identify familiar faces at a glance even after years of separation.”

(Turk & Pentland, 1991, p. 71)

Faces are much more than just useful keys to individual identity – the face is the focus of attention during social interaction and is also the part of the body that reveals an individual’s emotional state (Bruce, 1988). Face recognition is the most commonly used aspect of our abilities in remembering faces as we are required daily to remember or recognise a person or face. Face recall, however, is far less often used (Davies, Shepherd & Ellis, 1978; Laughery, Duval & Wogalter, 1986; Shepherd & Ellis, 1996). Occasions do however arise when we need to recall faces, such as a witness to a crime recalling what a suspect looked like and providing police with a description. The police make great use of verbal descriptions and other forms of face recall, such as artists’ sketches and face composites constructed by the witness or victim. Unfortunately, face recall is generally much more difficult than face recognition (Cohen & Nodine, 1978), which is possibly exacerbated by current composite techniques that require witnesses to decompose a holistic face into constituent parts or features, therefore interfering with the witness’s ability to maintain the image of the face (Ellis, 1986; Shepherd & Ellis, 1996). This individual feature processing is contrary to the normal facial coding process, as illustrated by many studies (Bruce, 1988; Davies, 1981; Davies & Christie, 1982; Ellis, Davies & Shepherd, 1978b; Laughery & Fowler, 1980; Wells & Hryciw, 1984).

The topic of facial memory, processing and recognition has been researched and analysed for almost a century. Various theories have been developed and

a multitude of findings have been reported, but nothing points to a definitive answer as yet. The central problems underlying the research are that current systems produce poor quality composites, especially when constructed from memory. Practical tools that enable witnesses of crime to recall and recognise the faces of perpetrators are relied upon by law enforcement organisations worldwide. The reconstruction of these faces is usually performed using a variety of methods, ranging from sketch artists to manual and computerised composite systems such as Identikit, Photofit, E-fit and Mac-a-Mug Pro (Tredoux, da Costa, Nunez & Rosenthal, 1999). Such face composite systems are based on facial recall, often making use of verbal descriptions as a starting point for initial feature selection (Shepherd & Ellis, 1996). Laughery and Fowler (1980) suggest one reason for the limitations of these face-recall techniques could be their reliance on the verbal descriptions. However, most of the literature examining the abilities of these techniques has shown poor results. Numerous studies have found that past and current composite-building techniques do not achieve accurate, recognisable composites of target faces (Christie & Ellis, 1981; Ellis, Davies & Shepherd, 1978a; Ellis, Shepherd & Davies, 1975; Green & Geiselman, 1989; Kovera, Penrod, Pappas & Thill, 1997; Laughery & Fowler, 1980; Prag, 2004).

Another key problem with existing composite systems is their reliance on featural composition and the lack of appropriate facial features contained in their databases. There is much evidence that faces are stored holistically rather than as lists of features, and extracting these features from the integrated form proves difficult for subjects (Davies & Christie, 1982). Davies and Christie (1982) suggest that featural-based composite systems, such as Identikit and Photofit, continually expose subjects to multiple sets of features during reconstruction, which may interfere with and degrade subjects' memory of the target face due to the incompatibility between processing strategy (holistic) and task demands (feature-by-feature in isolation).

The fundamental concept underlying all of these composite systems is the same - configural construction of a face by selecting features from an existing set (Kovera et al., 1997). The assumption that human visual recognition

operates on the principle of individual feature processing alone reveals the flawed theoretical basis upon which the systems were created (Brignull, 1998). Many studies have illustrated that facial patterns are processed as more than the sum of its independent parts, and that individual feature processing is contrary to the normal facial coding process (Bruce, 1988; Davies, 1981; Davies & Christie, 1982; Ellis, Davies & Shepherd, 1978; Laughery & Fowler, 1980; Tanaka & Farah, 1993; Wells & Hryciw, 1984; Young, Hellawell & Hay, 1987). This could account for the fact that there is no evidence that computerised systems lead to better reconstructions than mechanical systems.

Despite the disappointing empirical findings, composites are nevertheless invaluable to police inquests. Even with their shortcomings, they are useful for eliminating unlikely suspects in mug files or lineups prior to exposure to witnesses, narrowing the potential group of suspects and improving chances of identifying and apprehending offenders (Brignull, 1998; Davies & Valentine, 2004; Green & Geiselman, 1989; Prag, 2005).

Another central problem is the lack of a theoretical model regarding the mechanisms of facial processing and memory. Various theories have been formulated regarding the encoding and memory for familiar and unfamiliar faces, suspect attractiveness and perceived personality, and whether this processing strategy is purely featural, holistic or a combination of the two ('dual' processing), but there is disagreement amongst theorists. A consensus has not been reached, but most researchers favour a processing theory whose important elements are holistic.

Given the problems associated with producing accurate composites using the aforementioned facial reconstruction systems, this study tests a fairly new eigenface-based composite system called "ID" (Rosenthal, de Jager & Greene, 1998; Tredoux et al., 1999; Tredoux, 2001, Tredoux, Nunez, Oxtoby & Prag, 2006). This system is recognition-based, where the user chooses from a display of facial composites (called the graphical user interface or GUI) rather than constructing their own. This system was developed to compensate

for the numerous limitations inherent in current facial composite systems. It presents the face as a whole rather than fragmented features, which is consistent with the way people encode faces (Ellis et al., 1978; Laughery & Fowler, 1980; Wells & Hryciw, 1984). It also removes the need for verbal descriptions and expert operators.

This study examines whether ID, a holistic recognition-based composite system, provides improved identification accuracy over that of a featural recall-based composite system, namely Faces. McQuiston and Malpass (2000) examined the prevalence of composite systems used by police officers in the United States and the three most frequently reported systems used were the Identikit 2000, FACES and Comphotofit. Therefore the comparison aimed to test whether the ID system produces better quality composites than a computer-based system used by international law enforcement agencies including South Africa (FACES), as it is based on empirical evidence of people's superior ability to recognise a face rather than the recall of features in isolation, required by featural systems such as FACES.

The effect of reinstating contextual information on eyewitness identification performance using the guided memory interview has been examined in several studies (Cutler, Penrod & Martens, 1987; Krafka & Penrod, 1985; Malpass & Devine, 1981). These studies found that guided memory led to greater accuracy and showed improvement in recognition accuracy, proving that guided memory has a significant effect on memory and recognition performance. Therefore, this study aimed to assess whether participants given a guided memory interview prior to reconstruction would produce better quality composites than those participants given the standard treatment observed in reality, namely a police interview and whether this improvement is found for both featural and holistic systems.

In summary, this study is based on two key intellectual questions:

1. Do composites built using configural systems lead to better recognition than those built using featural-based systems?

2. Does context reinstatement sufficiently improve face memory so that recognition is better than when not given context reinstatement?

The search for a suitable technique to aid witnesses to remember a suspect's face and recognise that face later is an important motivating factor for this kind of research. Through such research, methods of composite construction can be reviewed and improved by eliminating aspects found to be ineffective and introducing aspects found to be useful. Interviewing techniques used by law enforcement can also be enhanced to improve eyewitness memory for the event and suspect.

Outline for this thesis

First, a review of the literature is discussed in Chapter 2, looking at the various methods of facial composite production, face processing theories and context reinstatement techniques.

The two composite systems are evaluated across various conditions for novice users after being exposed to a live staged event, and will be discussed in more depth in Chapters 3 and 4. The optimality of the two systems is also tested by expert operators who created in-view composites of the three perpetrators. Furthermore, this study looks at the effects of morphing composites from multiple witnesses. Motivation for this can be found in Bruce, Ness, Hancock, Newman and Rarity (2002), who found that combining face composites by morphing were rated as well as the best individual composite and significantly better than the average individual composite. This condition evaluates the practical ability of the system relative to a forensic context, where there are often multiple witnesses to a crime. The resulting individual composites often vary in terms of likeness, therefore the assumption is that combining these individual composites should produce a 'super composite', one that is a much better likeness to the perpetrator as it combines each of the witnesses' memories of the face. These and other results will be examined in chapter 4, with a discussion of the findings and general conclusions in Chapter 5.

Chapter 2

Review of the Literature

According to Clifford and Davies (1989) there are 3 phases where identification evidence is used by the police: *the descriptive phase*, where identity information regarding a suspect is obtained, either as a verbal description or composite or both; *the search phase*, where the obtained information is compared with police records and other witness recollections in an attempt to identify a suspect; and *the identification phase*, which involves the witness attempting to identify the suspect from a lineup. Most research has focused on evaluating the descriptive phase and is reported here.

1. Methods of facial composite production

Verbal descriptions

Verbal descriptions are the starting point in any criminal investigation. 3 types exist, namely free, cued and prompted description (Clifford & Davies, 1989). A *free* description involves the witness giving a description of the suspect without interference or prompting from the investigating officer. In a *cued* description, the witness responds to specific questions asked by the investigator, in the same way as the police interview used in this study (please refer to Chapter 3 for more information). A *prompted* description differs from the cued description in that it offers multiple answers to each question from which the witness must choose the most correct. Clifford and Davies (1989) suggest that the immediate, spontaneous statements made by witnesses should be considered as most important, and where further questioning is necessary, it is best to use cued descriptions rather than prompted descriptions that contain multiple-choice questions.

A standard method utilised by police to gather information about suspects is obtaining a verbal description of the suspect. Verbal descriptions are a primary means of communicating information about appearance, with or without the aid of a composite or photograph of the suspect (Shepherd, Davies & Ellis, 1978). Witnesses are required to complete a statement consisting of a list of details of the suspect's facial features, his clothing and jewellery and other particulars they can remember that will aid in the investigation. Chance and Goldstein (1976) found that participants asked to describe a target face recognised that face moderately better one week later than those participants asked to associate the face with something similar or provided no written response. However, a number of subsequent studies have demonstrated detrimental effects of verbal description on later identification of a target face (Brown & Lloyd-Jones, 2002; Dodson, Johnson & Schooler, 1997; Finger & Pezdek, 1999; Meissner & Brigham, 2001a; Schooler & Engstler-Schooler, 1990).

This phenomenon of verbal overshadowing, where recognition for faces could be impaired if the witness was required to verbally describe the perpetrator prior to recognition, has been extensively studied within the facial memory paradigm (Dodson et al., 1997; Finger & Pezdek, 1999). Schooler and Engstler-Schooler (1990) believed that those subjects who had not described the face based their recognition decisions on their visual memory, while those subjects required to make a description were biased to rely on their memory of that description, their verbal memory in effect overshadowing their visual memory. Cohen and Nodine (1978) also found evidence that faces may be encoded using visual rather than verbal imagery. The literature regarding verbal overshadowing proves that it is a genuine phenomenon that reliably occurs, however, due to the many empirical disagreements and varying results across such literature, it is difficult to formulate one conclusion regarding the mechanisms responsible for the overshadowing effect (Meissner & Brigham, 2001a).

A study by Christie and Ellis (1981) compared verbal descriptions with visual impressions (namely Photofit composites) and found that judges identified

targets significantly better when using the verbal descriptions than when viewing the face composites. When subjects were tested using both the Photofit composite and verbal description, correct identifications did not increase. This is counter-intuitive to general assumption: “surely a picture is worth 1, 000 words?” (Davies, 1983, p. 115). Christie and Ellis suggest that the difficulty lies in translating the verbal information into an accurate visual image; therefore the problem is not embedded in the witness’s verbal ability to describe a face but in some other aspect of the facial composite system. These ‘interference effects’ have now been widely accepted as a major problem with these composite systems and will be dealt with in more depth later in the thesis (Brignull, 1998) (refer to the section “Why are face recall systems so limited?” on page 22 for more information). However, Christie and Ellis’s finding of less than 50% identification accuracy when using the verbal descriptions does seem to suggest that people are generally not proficient at describing faces.

The assumption that people are generally not good at verbally describing faces is likely to be a limiting factor when constructing faces using techniques that rely on such descriptions as a basis. Human vocabulary seems to lack sufficient terms to characterise the general physiology of the face in enough detail and accuracy in order to define a single individual out of the general public. Due to this deficit, a description alone of a suspect’s face is unlikely to convey sufficient information in terms of identification (Brignull, 1998; Christie & Ellis, 1981; Laughery & Fowler, 1980; Shepherd et al., 1978).

Practical tools that enable witnesses of crime to recall and recognise the faces of perpetrators are relied upon by law enforcement organisations worldwide. The reconstruction of these faces is usually performed using a variety of methods, ranging from sketch artists to manual and computerised composite systems such as Identikit, Photofit, E-fit and Mac-a-Mug Pro (Tredoux et al., 1999). Such face composite systems are based on facial recall, often making use of verbal descriptions as a starting point for initial feature selection (Shepherd & Ellis, 1996). Laughery and Fowler (1980) suggest one reason for the limitations of these face-recall techniques could be their reliance on the

verbal descriptions. However, most of the literature examining the abilities of these techniques has shown poor results.

Sketch artistry

Sketch artistry is another alternative to using facial composite systems to capture a perpetrator of a crime. According to Osterburg and Ward (1997), the witness describes the perpetrator and answers any questions asked by the artist. The artist then works free from interruption and observation of the witness, using the information gathered from interviewing the witness. Upon completion, the first attempt is shown to the witness, who subsequently guides modifications by the artist. This process is repeated until the witness is satisfied or the artist believes the image cannot be improved (Osterburg & Ward, 1997).

Laughery and Fowler (1980) tested the effectiveness of the sketch artist and Identikit techniques, where 142 subjects worked with sketch artists and Identikit operators to build a sketch or composite from a description for 71 different target faces, replicating standard law-enforcement procedure. In addition, each artist and Identikit operator constructed another composite of each target with the face in view. They found that in both conditions, sketches were more easily identified than the Identikit composites. Laughery and Fowler propose that the Identikit system can create only a limited set of alternative faces due to its preset collection of features, whereas a sketch artist can possibly produce an infinite set, which may account for this outcome. At times the "perfect mouth" just may not be available among the features predefined in the Identikit archives. A second explanation for sketch superiority is that some kinds of detail such as shading and age lines are typically more prevalent in sketches than in composites. (This has subsequently improved over the years, with current composite systems containing shading, age lines and so forth of high quality). Another reason which could account for the outcome of sketch superiority is the total time spent generating sketches and composites. More time is spent producing sketches than composites; hence the witness spends more time thinking

about the target face which could possibly lead to more accurate descriptions and composites.

An advantage of the sketch artist approach is that the resulting image looks realistic and can be altered precisely according to the verbal description given by the witness. However, this method is restricted by the witness's verbal ability to describe the suspect, since every aspect of the suspect's face has to be described spontaneously from memory. Another limitation is that, practically, not every police officer has the ability or talent to be a sketch artist and cannot be trained to do so, therefore specialist artists need to be employed. It is also a very time-consuming process (Brignull, 1998).

Sketch artistry was the first practical method of facial reconstruction, but was limited by the witnesses' ability to describe the target face and the artist's ability to accurately sketch the features being described (Kovera et al., 1997). It was also the only method of facial reconstruction until the development of facial composite construction systems. These manual or mechanical composite systems consist of numerous facial features, which need to be united to form a whole face. Two of the most well known systems that have been widely used across the globe are the Identikit and Photofit systems (Kovera et al., 1997).

Manual and computerised systems

MANUAL SYSTEMS

Identikit

Hugh MacDonald, a Californian police officer recognised the need for a uniform system that could reproduce face likenesses without the involvement of a skilled sketch artist, and subsequently introduced the Identikit in 1959 in the United States (Davies, Ellis & Shepherd, 1985; Davies & Valentine, 2007). The Identikit consists of numerous line drawings of facial features printed on transparent sheets, which are superimposed to create a composite. The

system contains five sets of facial features, namely hairstyles, eyes, noses, mouths and chins, as well as a variety of accessories such as hats, glasses, facial hair and age lines. A marking pencil can also be used to further modify the composite (Barber, 1988; Brignull, 1998; Shepherd & Ellis, 1996).

The Identikit has been criticised for its limited effectiveness in producing composites. As mentioned earlier, Laughery and Fowler, (1980) discovered that sketches made by sketch artists were judged more accurate than Identikit composites. It has also been criticised for its limited ability to represent salient or distinctive faces. Green and Geiselman (1989) tested Identikit composites of faces with and without a distinctive feature and found that more accurate identifications were made for Identikits without a salient feature than those with a salient feature. This is contrary to findings in face recognition, where salient faces are better remembered and recognised than typical faces (Light, Kayra-Stuart & Hollander, 1979). Green and Geiselman believed their results were due to the difficulty of finding features in Identikit to match the distinctive feature of the target face.

Many studies have shown the inadequate flexibility of the system and its limiting technique in representation accuracy, and together have led to the conclusion that Identikit generally produces poor composites, which bear little resemblance to the faces they are meant to represent (Kovera et al., 1997).

Photofit

The British equivalent to the Identikit was developed shortly after by Jacques Penry in 1970 (Shepherd & Ellis, 1996). Photofit is similar to Identikit, and consists of black and white photographed features printed on cards that can be inserted into a special frame. As with Identikit, there are five sets of features, namely forehead-hairline, eyes, nose, mouth and chin, with similar features (such as thick lips or large eyes) grouped together. Transparencies can be placed over the frame to add details (Barber, 1988; Kovera et al., 1997).

Photofit has demonstrated similar results to those of Identikit studies (Christie & Ellis, 1981; Davies, Ellis & Christie, 1981; Ellis, Davies & Shepherd, 1978a; Ellis, Shepherd & Davies, 1975). Essentially, studies have shown that Photofit constructions do not produce reliable and accurate composites.

Ellis, Shepherd and Davies (1975) asked subjects to reproduce, using Photofit, a composite constructed from randomly selected features of the Photofit kit. Subjects had difficulty reconstructing the Photofit faces from memory, even when the target face was present and all the necessary features were available. Subjects found it difficult to select the correct set of features from the Photofit database. In a second experiment, they used real, photographed faces which were reconstructed from memory. Correct identifications by judges were only 12.5% (1 in 8), which was above chance levels, but disappointingly low.

In a second study, Ellis, Davies and Shepherd (1978a) further investigated some of the factors affecting the quality of Photofit constructions. They found that reconstructions gave poor impressions of the target faces, whether made from memory or while the face was in view, that composites of a live target were no better than for a photographed target, and instructions to remember the target did not enhance the witness's performance, leading to the conclusion that there are inherent limitations in the efficiency of the system. In fact, the results showed only a slight advantage in using Photofit over witnesses sketching the face from memory themselves! Davies, Ellis & Shepherd (1978) noted similar limitations of the Photofit system.

As mentioned earlier, Christie and Ellis (1981) found that verbal descriptions of a target were significantly more accurate than Photofit composites, contrary to the general assumption that people are not good at verbally describing faces (Ellis et al., 1978a; Laughery & Fowler, 1980). The overall conclusion is that the Photofit system does not create accurate, recognisable facial representations and that it is limited in its system design (Laughery & Fowler, 1980).

In order to assess whether real witnesses to crimes would do better than research subjects, the British Home Office conducted a systematic survey of the operational effectiveness of Photofit (Kitson, Darnbrough & Shields, 1978). Follow-up questionnaires were circulated to investigating officers where Photofit composites were made during police enquiries. Of the 140 cases that had been successfully solved, officers estimated that in about 5% the Photofits had been entirely responsible for solving the case, with a further 17% being very useful in solving the crime. However, in 45% of cases, the Photofits were not very useful or of no use at all (Clifford & Davies, 1989). Essentially, the survey suggests that experimental findings that Photofit produces poor composites are generally demonstrated in real-world investigations too. A later survey of Photofit composites constructed by the Metropolitan Police gave similar results (Bennett, 1986). Only 3.8% of cases where a Photofit was constructed had been solved, however of those solved cases, Photofit had been helpful in 59%. Unfortunately the usefulness of the composites cannot be generalised to the other 346 cases that were not solved.

One shortcoming is the range and representativeness of features in the kits. In an early survey of the Photofit's usefulness by King (1971), police respondents suggested that the range of features contained in the kit could be improved. In a later survey, Kitson et al. (1978) again found a need for a wider range of features in the kit. Other studies support this suggestion, for example "there may be times (and according to technicians, there are) when 'the right nose is not there'." (Laughery & Fowler, 1980, p. 313). Bennett (1986) notes that the equipment was in serious need of updating as there was a distinct lack of modern hairstyles and young facial features. It seems impossible for the composite systems to represent the infinite variations of features seen in real faces.

Another difficulty inherent in such manual systems is the underlying principle which assumes that people remember faces as component features. The ability to manipulate the face globally and to store larger, more representative feature sets required discarding the old mechanical systems for more versatile

and powerful technology, namely the modern computer (Davies & Valentine, 2007).

COMPUTERIZED SYSTEMS

Following the results reported above, computerised composite systems such as Mac-a-Mug Pro and E-fit have been devised to compensate for the inadequacies of past systems. The computer-driven systems offer improvement over manual systems as they offer a much wider range of facial features and are able to manipulate these features with the assistance of graphic packages. However, despite the computerised systems' advantage over manual systems, they still do not achieve accurate, recognisable composites of target faces (Prag, 2005).

Whatsisface and CADC

Early attempts to computerise manual composite systems resulted in the creation of systems such as WHATSISFACE and the CADC System. WHATSISFACE was the world's first computerised face compositor where a non-artist could create a facial image on a graphical display without the talent and intuitive knowledge possessed by a sketch artist (Gillenson & Chandrasekaran, 1975; Laughery, Rhodes & Batten, 1981). WHATSISFACE images were compared to sketches made by subjects with a photograph of the target face in view. The WHATSISFACE system was able to produce impressive reconstructions when done with the face in-view that were significantly better identified than the sketches, but gave poor quality results when reconstructed from memory.

Gillenson and Chandrasekaran (1975) uncovered the potential of computer graphics in providing an extremely versatile composite tool. The British Home Office in conjunction with the Computer-Aided Design Centre (CADC) in Cambridge developed a prototype system (Christie, Davies, Shepherd & Ellis, 1981). Users were able to alter the size, shape and position of any feature as well as distort the entire face, essentially allowing unlimited number of

possible faces to be produced. The join lines normally visible in Photofit composites were also eliminated by a blending function, producing a more life-like face. Christie et al. (1981) compared the likenesses of the CADC and Photofit systems, and found no difference between them on any of the evaluative measures for quality of likeness. Both systems produced more recognisable composites when done with the target face present compared to its absence, however CADC composites made from memory were marginally better than the Photofits. Nonetheless the results were disappointing with an overall identification accuracy of only 30%.

The 1980's introduced the arrival of the desktop computer, along with a rapid development in information technology. This period showed a considerable drop in cost of hardware and graphics packages, which allowed police forces for the first time to use computers extensively. During this period numerous computerised facial composite systems entered the market, such as E-fit, EvoFit and Mac-a-Mug Pro (Brignull, 1999; Davies & Valentine, 2007; Laughery et al., 1981).

These new computerised systems have some benefits over the original manual systems, including a significantly expanded range of features as well as the ability to manipulate these features through graphics packages (Davies, van der Willik & Morrison, 2000). These computerised systems also offer flexibility beyond that of the older manual ones, for example, independent feature movement and manipulation, the use of editing tools to build a composite that fits more closely with a witness's description, and the ability to export composites to other applications such as Adobe Photoshop where they can be further edited (Brignull, 1998). Operators can now move features independently, adjust the dimensions of individual features and the face as a whole, and manipulate the composite in ways that previously only sketch artists were capable. Two of the most well known and extensively researched computerised composite systems are Mac-a-Mug Pro and E-fit.

Mac-a-Mug Pro

Mac-a-Mug Pro is a line-drawing computer system designed for use on a Macintosh computer and was developed from the original Identikit. It comprises of an expanded database of line-drawn facial features which can be assembled and displayed as a face on a display screen, whereupon further modifications can take place with the aid of specialised editing programmes. Mac-a-Mug Pro can produce potentially unlimited combinations of features due to these editing features. For example, features can be enlarged or shrunk, features can be moved around independently, and unique marks such as age lines and scars can be added (Davies et al., 2000; Davies & Valentine, 2007; Shepherd & Ellis, 1996).

Photofit, Identikit and Mac-a-Mug Pro have been most extensively researched of all the composite systems. Of these three systems, Mac-a-Mug Pro seems to have the most potential because of its large database of stored features and its flexibility during editing (Koehn & Fisher, 1996).

Research by Cutler, Stocklein and Penrod (1988) confirmed the system's potential to build recognisable and realistic composites. The authors compared identification rates of subjects using photographs and Mac-a-Mug Pro composites created by an experienced operator as references when searching for the target in a mug-shot file. They found that the expert composites were as effective as the photographs in aiding identifications. This study demonstrates that under optimal conditions Mac-a-Mug Pro is capable of producing very recognisable composites.

Wolgater and Marwitz (1991) asked participants to create composites from memory and with a target photograph in view. They investigated whether the face composites constructed from memory and with a photograph present would differ in quality, and found that composites produced while viewing the target image were better than the from memory composites, with judges matching the composites to original photographs of the targets at above chance levels. However, studies of more forensic realism, that are performed

under similar conditions relevant to (realistic) forensic settings, have produced more pessimistic findings (Davies et al., 2000).

In a study by Koehn and Fisher (1996), participants interacted with a stranger before being asked to reconstruct his face from memory using Mac-a-Mug Pro. The resulting composites were rated for degree of likeness to the target. Overall likeness ratings were very poor – “the composites were rated as being not even remotely similar to a photograph of the target person” (p. 221) – with only 4% correct identifications from the photo spread. The notion that good quality composites can be created when the target is present but not when the composite is created from memory is supported in the follow-up study, where ‘ideal’ composites made by an experienced operator with the target face present were correctly identified by 77% of the judges.

These pessimistic findings were reiterated by a series of experiments conducted by Kovera et al. (1997). Study 1 made use of familiar faces as targets, rather than strangers. Students were required to reconstruct faces of former classmates and teachers from their high school. These composites were then shown to fellow classmates familiar with the targets, who acted as judges in evaluating composite quality. They could not differentiate the familiar (old classmates and teachers) from the unfamiliar composites (new faces), even though they were aware that the composites were of faces familiar to them. Judges rarely generated a name for a composite: only 3 in 167 names (1.7%) offered for the composites were correct. In light of these findings, it seems that the Mac-a-Mug system is severely limited in producing recognisable composites from memory under more forensically realistic conditions.

Study 2 asked naïve participants to determine whether or not they had previously seen a composite of one of the faces in a lineup. They could not consistently select the people depicted in the composites from the photo lineups consisting of the target photograph and four foil photographs. The results from these studies raise doubts about whether people who view such composites will be capable of identifying people they encounter in everyday contexts as the suspect depicted in the composite (Kovera et al., 1997).

Considering that face composites are always done from the witness's memory in typical eyewitness situations, Koehn and Fisher (1996) and Kovera et al. (1997) believed that the Mac-a-Mug Pro system was not useful for real-world investigations, where witnesses must construct the composites from memory. The poor performance in memory conditions may be due to the nature of the construction task, namely piecemeal assembly of features, which is incompatible with the configural face processing procedure. Christie et al. (1981) and Davies, Ellis and Shepherd (1978) found that feature-based reconstruction techniques interfered with later recognition and only gave a facial type likeness. This individual feature processing is contrary to the normal facial coding process, as illustrated by many studies (Bruce, 1988; Davies, 1981; Davies & Christie, 1982; Ellis, Davies & Shepherd, 1978a; Laughery & Fowler, 1980; Wells & Hryciw, 1984).

A proposed solution was to start with one of a limited set of prototype faces whose feature components can be changed within the context of the face (Davies & Christie, 1982). Koehn and Fisher (1996) and Kovera et al. (1997) also speculated that a composite system that encourages a more holistic, configural approach to face construction may give better quality likenesses.

E-fit

Such a system is the E-fit (Electronic Facial Identification Technology), a computer-based system that makes use of photographic features stored in a feature library, much like Photofit (Davies et al., 2000; Davies & Valentine, 2007). The designers of E-fit wanted to create a system which was more attuned to existing knowledge from face perception research, and which did not require witnesses to select the features of the face in isolation. For this reason, E-Fit was designed to present features in the context of a complete face. The user is prompted via on-screen menus to provide verbal descriptions of the various features, from which a search protocol selects the most appropriate features from the database, and an initial impression is displayed. The user can then amend this by searching through the library's array of alternative features and substituting individual features or adjusting

the size or configuration of features within the context of the face. The consequence of a featural change is immediately obvious in a holistic sense as it is viewed in the context of face, however this selection process itself forces users to recognise disembodied features from the library before viewing them contextually.

Davies et al. (2000) compared the mechanical Photofit system with the computer-driven E-fit system in constructing familiar and unfamiliar faces. Target faces were first reconstructed from memory and then with photographs in view. Composites were assessed by judges familiar with the target faces and who were in daily contact with them. They rated composites for familiarity, attempted to name facial composites where possible and also attempted to match composites to the target photographs. Results revealed that E-fit was consistently superior when a familiar face was constructed in the presence of photographs (83% matching accuracy) and was also quicker in reproducing faces. However, the superiority of E-fit is confined to target-present conditions, as E-fit was no better than Photofit when the composites were made from memory. Naming was also problematic: correct naming only occurred for composites of familiar targets that were constructed with the target present. These findings are consistent with earlier findings by Kovera et al. (1997). Results indicate that E-fit is not consistently superior in producing facial likenesses under laboratory conditions to the old mechanical systems like Photofit – a peculiar result given that E-FIT was developed to overcome Photofit's deficiencies (Frowd, Carson, Ness, Richardson, et al., 2005).

Koehn (1995) was the first to make a direct comparison between the E-fit and Mac-a-Mug Pro systems. Koehn hypothesised that the sophistication of the E-fit system would produce better quality composites than the Mac-a-Mug Pro. Participants viewed a video-clip and returned the following day to give an in-depth verbal description or construct a composite using either the E-fit or Mac-a-Mug Pro system. The comparison demonstrated that E-Fit composites were of better quality than Mac-a-Mug Pro composites, however composites from both systems were found to be useless for identifying the target person from a lineup. Additionally, lineup performance was so low that both E-Fit and Mac-a-

Mug Pro composites were no more useful than verbal descriptions (Koehn, 1995).

Further low results were found by Davies and Oldman (1999). They explored the effect of positive or negative attitude towards a target on future recognition of an E-fit reconstruction. They used celebrity faces as targets, about whom subjects were known to hold positive or negative views. Composites were constructed from memory and with a photograph of the target present. Subjects correctly named only 7.6% of the targets overall, which rose to 12.5% when composites were made from photographs by those who disliked them. They found that attitude towards the target did significantly influence accuracy of reconstructions, with ranking and naming data showing that the best composites were made by subjects who disliked the target and made their reconstructions with a photograph present. In light of their results, Davies and Oldman believed that negative feelings toward a suspect are likely to aid rather than hinder the construction of a likeness, which has important implications for the general quality of composites made by witnesses to actual crimes, who must construct composites of people for whom they have intense negative feelings.

More favourable results were found by Brace, Pike and Kemp (2000). Most previous studies have used unfamiliar target faces during testing. In reality, composites are created in the hope that the image will be seen and recognised by someone who is familiar with the suspect. Brace et al. decided to use famous faces as targets in order to test the efficiency of the E-fit system. Composites constructed with a photograph of the target in view were compared to those constructed from memory. There were two modes of construction condition, namely whether the E-fit was constructed with a witness describer (who viewed a photograph of the target and described the face to the operator) or directly by the E-fit operator. Significantly more E-fits were correctly identified when constructed by the operator (34.72%) compared with those constructed with the witness describer (24.95%). Consistent with the findings of Davies et al. (2000), it was found that the superiority of E-fit is confined to target-present conditions.

The assumption that human visual recognition operates on the principle of individual feature processing alone reveals the flawed theoretical basis upon which the systems were created (Brignull, 1998). There is much evidence that faces are stored holistically rather than as lists of features, and extracting these features from the integrated form proves difficult for subjects (Davies & Christie, 1982). Tanaka and Farah (1993) found that subjects were better at recognising features of a target face when embedded within the whole face than when these features were presented in isolation, therefore implying a holistic processing of faces. Featural-based composite systems continuously expose subjects to multiple sets of features during reconstruction, which interferes with and degrades subjects' memory of the target face due to the incompatibility between processing strategy (holistic) and task demands (feature-by-feature in isolation) (Davies & Christie, 1982). This could account for the fact that there is no evidence that computerised systems lead to better reconstructions than mechanical systems.

Facial composite systems have been shown to perform unacceptably poorly in a variety of studies over many instances, and the basis of their poor performance has generated a great deal of research focused on the key question: "Why are face recall systems so limited?" (Brignull, 1998).

2. Why are face recall systems so limited?

Five possible reasons emerge from the literature (Barber, 1988; Brignull, 1998; Davies, 1981; Davies & Valentine, 2007; Rosenthal, 1998; Shepherd & Ellis, 1996):

- Firstly, the way composites are judged is unfair to the systems, namely judges tend to rate composites on the lower end of the scale in rating, sorting and identification tasks commonly used to measure composite accuracy. People do not realise that the composites are created from memory usually after some delay, and are meant to be likenesses only

and not an exact resemblance to the target face, and subsequently rate most composites poorly.

Attempts have been made to increase positive identifications with these manual composite systems by improving the quality of composites. Ellis et al. (1978b) found that the feature demarcation lines present in Photofit constructions reduced identification rates in recognition tasks, and possibly interfered with face processing. The removal of these lines was hoped to improve the quality of Photofit constructions. In another study, Gibling and Bennett (1994) investigated whether artistic enhancement of a basic Photofit composite by a skilled operator would improve the quality of the Photofit produced and hence improve identification rates. Composites that had been enhanced gave significantly more correct identifications than the basic non-enhanced ones. These studies show that there is perhaps potential for manual systems if enhancement techniques are used.

- Secondly, the limited expressive capabilities of the systems to generate good quality composites, mostly due to the limited range and representation of features available, as well as an inability to represent faces that are atypical or distinct as mentioned earlier by Green and Geiselman (1989).
- Thirdly, having to attend to different features in isolation and the selection of such features out of context of the face interferes with the memory of the face being constructed. Specifically, it has been found that judging likeness from isolated features may seriously interfere in the process of constructing composites and the accuracy thereof (Davies & Christie, 1982). The human visual recall system is not perfect – while witnesses are able to recall major, cardinal features (those that are recalled most frequently, such as hair), they are unable to recall specific details of someone seen briefly, especially after a substantial delay (Shepherd & Ellis, 1996).
- Fourthly, faces are not remembered as a collection of isolated features and therefore involve an integrated representation of the face. Methods heavily dependent on scrutinizing features in isolation will hence be

poor. Judgment of features is much more accurate when made in the context of a whole face than in isolation (Davies & Christie, 1982), thus a more holistic approach to face recall would potentially provide better results (Laughery & Fowler, 1980; Davies & Christie, 1982).

- Lastly, the initial stage of such systems entails a verbal description of the target face given to an operator who acts as an intermediary between the witness and the system. The witness must articulate the mental image to the operator, who in turn translates the description into a facial image. This transfer of information is very susceptible to corruption and causes much of the difficulty in generating likenesses (Kovera et al., 1997; Laughery & Fowler, 1980; Wolgater & Marwitz, 1991). In addition, the poor conformity of reconstructions relative to the target face could be due to the limited ability of the witness to provide an accurate verbal description of the face, or that verbalisation interferes with the visualisation process relied upon by face recall.

Elaborating on the fourth point made above, much theory has been generated around the different face processing strategies used for memory of faces. Three theories emerge – featural, holistic and dual – and each will be described next.

3. Face processing theories

Featural face processing

Bradshaw and Wallace (1971) presented pairs of Identikit faces differing from each other with respect to two, four or all seven features, and asked subjects to make same-different judgements. They hypothesised that the number of differing features should not affect reaction time if the faces are processed in parallel or holistically. They found a systematic decrease in reaction time as the number of differing features increased – participants judged the faces to be different quicker as more differing features were inserted – therefore

affirming that faces are treated as a collection of independent features and are not processed as a whole.

Leder and Bruce (2000) challenged the importance of facial context in face recognition. They argued that isolated featural information plays an important role in face recognition, and that its processing does not require the context of the whole face (holistic processing). They found a remarkable ability of subjects to recognize isolated relational information, and there was no significant increase when context was added to encourage holistic processing. Their results support the notion that facial information is processed in an independent, featural manner.

During the course of the investigation into featural face processing, some researchers discovered that certain features of the face were more significant than others. Shepherd et al. (1981) measured the frequency of feature mentioning. They found that aspects of the upper face were paid more attention than the lower face. The relative importance of the upper region of the face is also suggested by Fisher and Cox (1975, in Shepherd et al., 1981), who found that faces were harder to recognise if the upper part of the face was hidden. The multidimensional scaling studies reviewed by Shepherd et al. (1981) showed that hair and face shape were the most salient features.

Haig (1986) examined feature saliency using a computer system that allowed features from one face to be stretched, compressed, deleted or transferred to another face. Participants were asked whether components of a face were changed from study to test. Haig found that participants were best at noticing a change of the face outline, followed by the eyes, then the mouth, and lastly the nose. Changes to features in the upper half of the face (i.e. hair and eyes) were recognised better than changes to features in the lower half of the face (i.e. nose, mouth and chin). McKelvie (1976) also reports that the eyes are more important than the mouth in face recognition, as he found that covering the eyes of a learned face led to decreased recognition due to a loss in facial expression (primarily conveyed by the eyes).

Besides upper versus lower feature regions, inner (central) versus outer (peripheral) feature regions have also been contrasted to assess their ability to serve as cues for recognition. The importance of these categories of features seems to be dependent on familiarity. Ellis, Shepherd and Davies (1979) found an advantage for identification for inner features when recognising famous faces, while no difference was found in recognition rates when given inner or outer features in recognising unfamiliar faces.

In addition to feature saliency, some experiments have also found an order of face processing. Walker-Smith, Gale and Findley (1977) monitored eye movements while facial pictures were being viewed and found a top-to-bottom order of processing. Hines and Braun (1990) measured order of face processing for familiar and unfamiliar faces using a same-different matching task. They found a significant top-to-bottom order of processing for unfamiliar faces but not for familiar faces, the processing of which may involve a parallel (or holistic) processing.

Holistic face processing

There is much evidence that faces are stored holistically rather than as lists of features. Tanaka and Farah (1993) found that subjects were better at recognising features of a target face when embedded within the whole face (73%) than when these features were presented in isolation (62%), therefore implying a holistic processing of faces.

Davies and Christie (1982) tested whether subjects could make accurate similarity judgements when features were not embedded in a face. They found that when a target face was present, similarity scores were high whether features were presented in isolation or as part of a face, but when the target face was absent, scores were high only when features were presented in the context of a face and not in isolation. They concluded that faces are stored in an integrated form from which it is problematic to extract individual features.

A study by Homa, Haver and Schwartz (1976) also found superior performance at identifying individual features when embedded in the context of a face. Subjects were asked to indicate which features had been present on a previously seen face. These faces were either intact or scrambled. Homa et al. found superior performance at identifying individual features when embedded in the context of an intact face than in scrambled form. They believed that a face can act as a perceptual gestalt in facilitating processing of facial parts.

Young et al. (1987) cut celebrity faces horizontally into upper and lower halves and recombined halves from different faces to form a new, unfamiliar face. Participants made recognition judgments based on the top half of each face presented, and results showed that they were quite accurate at identifying the top half of the face when it was seen on its own or when the two halves were off-set so that the face outline was disrupted, but when it was combined and properly aligned with the wrong lower half it became extremely difficult to recognise the upper features. Hole (1994) used a similar procedure in his study to show that the finding made by Young et al. (1987) can be applied to unfamiliar as well as to familiar faces. The fabricated faces (those made up of two different halves) appeared to induce strong unconscious configurational processing which interfered with the selective featural processing required in the tasks. However, when these fabricated faces were inverted, subjects found it easier to name the halves.

An intriguing experiment done by Thompson (1980) lends support for holistic face processing with his "Thatcher Illusion". Thompson took a picture of Margaret Thatcher, and cut out and inverted the eyes and mouth within the face. When viewed upside down, the picture looks quite normal and seems similar to the original photograph, but when viewed upright the two faces are noticeably different and the created picture looks quite grotesque. This illusion provides evidence that there are two face processing paths, and that configural processing is dominant when faces are viewed upright while featural processing is dominant when faces are inverted (Brignull, 1998).

Wells and Hryciw (1984) found that trait judgements about a face improved recognition memory compared to when featural judgements were made. They proposed that these trait judgements foster a holistic processing of a face, which is consistent with the normal facial coding process. This opinion was further explored by Shapiro & Penrod (1986), who performed a meta-analysis of 128 eyewitness identification and face recognition studies. It was found that the hit rate was significantly greater if the instructions orientated a subject to encode a face with a personality trait (holistic encoding) rather than to locate a facial feature (a feature-based encoding).

Each of these earlier lines of research was relevant to testing the hypothesis that faces are stored in a relatively holistic form in memory. However, the hypothesis that faces are *perceived* holistically has not been tested directly. That was the goal of Farah, Wilson, Drain and Tanaka's (1998) experiments. Participants made same-different judgments of pairs of faces that were either identical or differed by either just the named part or all parts. The compatibility of the sought after parts and irrelevant parts (those parts that were not sought) had an effect on response accuracy and the effect was larger for upright faces than for inverted faces. The results confirmed their hypothesis that face perception is holistic.

In a recognition task, holistic processing is required as the witnesses search through many whole faces until a response of familiarity is evoked. Face recall, however, entails a witness decomposing an initially holistically encoded image into its constituent parts and using these parts (or features) as references to seek the image associated with it. Needless to say that face recall is generally much more difficult than face recognition (Shepherd & Ellis, 1996).

Dual face processing

Coinciding with the vast research conducted on holistic face processing, there is also evidence for a dual processing theory that purports that faces are

perceived both featurally *and* holistically. Matthews (1978) was one of the first researchers to arrive at this conclusion. He presented pairs of Identikit composites differing from each other with respect to one, two, four, five or all six features, and asked subjects to make same-different judgements. Results showed that features of the subset hairline, eyes and chin were given processing priority and processed in parallel, followed by a slower feature-by-feature comparison of the remaining features (namely the eyebrows, nose and mouth), suggesting a dual processing strategy.

Tanaka and Sengco (1997) tested recognition of features shown in isolation, in a new face configuration (eyes were moved further apart or closer together) or the original configuration. Results showed that subjects recognised features best when presented in the original configuration (77%), followed by the new configuration (72%), and poorest in isolation (65%). They also found that altering one feature (the spatial location of the eyes) not only impaired subjects' recognition of that feature, but also disrupted their memory for the other features (nose and mouth), emphasising an interdependency of featural and holistic information in face recognition.

Further support for the dual face processing strategy can be found in Sergent (1984). She used a matching task with pairs of Photofit faces where three characteristics of the face were manipulated. Inverted faces showed a featural processing strategy while upright faces showed a configural, holistic processing. Results suggest that faces have both featural and configural properties, and these different processing strategies are not mutually exclusive and can occur simultaneously.

In an effort to measure contributions of featural and configural processing to face recognition, Cabeza and Kato (2000) constructed configural prototypical faces and featural prototypical faces. For each set of four faces, a "configural prototype" was created by morphing the four individual faces into one composite, as well as a "feature prototype" where a single composite face was made by combining the features of the four individual faces. Participants

tended to commit false alarms for nonstudied prototypes, and this tendency was equivalent for both featural and configural prototypes. Moreover, participants that viewed inverted faces committed fewer false alarms with the configural prototypes than the featural prototypes. These results suggest that both featural and configural processing are important in face recognition, and that their effects are inseparable, hence endorsing a dual processing strategy.

Rakover (2002) reiterates this sentiment, reporting that overall results of the numerous studies conducted in this line of research support the hypothesis that in face perception and remembering, featural, configural and holistic information (a combination of featural and configural processing) are of great importance, and these results cannot be explained by featural or configural information alone.

The outcomes of the above studies suggest that faces are not perceived strictly and exclusively as features or as a whole, but that the two modes of processing operate interactively rather than independently (Bruce, 1988; Rakover, 2002). Laughery et al. (1986) also conclude that people may process a face holistically or featurally, depending on the reason for memorising such facial information and the subsequent use thereof. "An initial holistic (or configural) processing may occur so that overall information about a face can be obtained quickly. This holistic processing then directs the subsequent featural processing of more detailed aspects of a face, and the two processes continue in parallel to yield more information about a face" (Bruce, 1988, p. 39).

A solution is being developed

Computerised composite systems allow much greater control over the manipulation of the properties of a face than was possible with the manual systems, and have seemingly overcome the problems of range and representativeness of features. These systems seem to have advantages over the manual composite construction methods, yet they still depend on

many of the same assumptions about how faces are stored and retrieved from memory (Clifford & Davies, 1989).

Existing systems appear to be based on a logical rather than a psychological analysis of the process of face composition, and it is essential that the systems be more closely aligned with the dynamics of facial perception and memory as revealed by experimental research for any major improvements to be likely (Clifford & Davies, 1989; Davies et al., 1985).

Recognisable composites can be built by these systems, provided they are created with a reference photograph in view. However, the problem still remains that these systems do not produce good likenesses when constructions are done from memory (Davies & Valentine, 2007). Part of the problem with these systems – whether manual or computerised – is they involve a single facial representation that requires a verbal interaction (Frowd, 2001). A growing body of research by computer scientists and engineers has explicitly addressed the issue of perceptual processing of face patterns in the search for artificial face recognition systems suitable for security and forensic applications (Hancock, Burton & Bruce, 1996). When a multiple face presentation is considered, people's innate ability to recognise faces is capitalised, as opposed to recalling of individual facial features, and the disruptive effect of the verbal component is diminished. This forms the basis of a new generation of composite systems, which endeavour to evolve a facial image within a face space (Davies & Valentine, 2007).

4. New generation configural composite systems: “Evolving faces”

EvoFit

This motivated the development of EvoFIT, a system that has moved away from the traditional feature-based approach to a more holistic-based approach. Frowd (2001) believes that a system such as EvoFIT, which

presents the user with 18 complete faces simultaneously, allows people to benefit from our face recognition ability. Users make a selection from the multiple randomly generated faces, according to whom they think represents the face of the suspect. Evolutionary algorithms are then used to “breed” these choices together to produce a new set of 18 faces that more closely resemble the target. During this breeding process the faces become more and more similar to each other and also to the target. This process continues until an acceptable likeness is achieved (Frowd, 2001). Therefore, a composite is created by ‘evolution’ (Frowd, Carson, Ness, Richardson, et al. 2005). Software tools, such as the Feature Shifter and Facial Composite Tool, are also provided which allow the size and location of any features to be manipulated at the user’s request. Frowd and various colleagues have done extensive research with the EvoFIT system and some of their findings are reported next.

In Frowd’s (2001) testing of the system, subjects were asked to construct a face composite of a famous person from memory using either EvoFIT or E-fit. To achieve a degree of similarity with real life situations, a Cognitive Interview-based approach was used after exposure to the target to obtain a verbal description of the face. In addition the identity of the targets was hidden from the two operators, since operators in a forensic context have no prior exposure to the suspect. Results demonstrated a spontaneous naming rate of 9.6% for EvoFITs, which was about 7% less than the E-fits constructed under the same conditions (17.1%). The age of the celebrities was found to be a factor that contributed to the poor performance of EvoFIT. In a follow-up study, spontaneous naming rates rose to 25.3% during the construction of more appropriately aged famous faces with the target present. The data compares favourably with Brace et al. (2000) who found a naming rate of 25% with E-fits of famous faces also constructed with the target present. This study shows that EvoFIT is capable of producing recognizable composites, but yet again only when the target photograph was present during construction.

Frowd, Hancock and Carson (2004) tested EvoFIT under more realistic conditions with age-appropriate targets. Mock witnesses viewed an unfamiliar

face for 1 minute. They returned 2 days later to build a composite from memory using either E-fit, PROfit or EvoFIT after being given a Cognitive Interview. Naming rates were very low – only 3.6% of EvoFITs and 1.3% of PROfit composites were correctly named, with no correct naming for E-fits – which is considerably less than the naming rates found by Frowd (2001). Results suggest that the longer time interval from witnessing the face to reconstruction produced very poor composites. Even an improved and more accurate EvoFIT face model, tested in a subsequent experiment, once again gave very low naming rates, but EvoFIT (8.5%) was significantly higher than PROfit (3.7%). Results show that the latest version of software does produce better composites, however improvements and further work on the system is still necessary to support face selection. Possible improvements suggested by the authors include increasing the number and the variety of faces in the face model, and increasing the variability of the faces so that they do not look too similar to each other (which cause difficulty in face selection) (Frowd et al., 2004).

Frowd, Carson, Ness, Richardson, et al. (2005) compared 5 composite systems (E-fit, PROfit, Sketch artist, Photofit and EvoFIT) under realistic conditions. Mock witnesses inspected a photograph of an unknown celebrity for 1 minute. Following a 3–4 hour delay, witnesses described the face using a Cognitive Interview and then worked with an experienced operator to construct a composite from memory using one of the 5 techniques. The highest naming rates were found for E-fit (19.0%) and PROfit (17.0%), followed by Sketch (9.2%), Photofit (6.2%) and EvoFIT (1.5%). This poor naming rate for EvoFIT was considerably less than reported previously (about 10% in Frowd, 2001) in spite of a more appropriately aged target set, which was believed to be a reason for the poor results previously discovered. In real life, composites are usually accompanied by context information including a description of the suspect, information about the crime scene and any other significant or relevant information, which may serve to elevate the naming rate (but perhaps also the number of false alarms).

Frowd, Bruce, McIntyre, et al. (2006) developed a set of psychologically useful scales that allow EvoFIT faces to be manipulated. Ratings were collected for each face along six holistic dimensions: attractiveness, health, honesty, extroversion, threatening, and masculinity. To explore the effectiveness of the new dimensions, three main evaluations were conducted using perceptual and identification tasks. Results show that a set of psychologically useful dimensions can successfully enhance composites produced from EvoFIT. When the original composite of each target was presented along with the manipulated composite and the target face, the manipulated image was preferred 75% of the time, significantly more often than the original composites. When asked to name the famous person depicted in the original or manipulated composites, the manipulated composites were correctly named 9.6% of the time, compared to 4.8% for the original composites, again a significant increase. Testing of the Holistic Tool developed for use by witnesses again showed that the manipulated composites were more often selected as the best likeness. Therefore, the holistic dimensions, implemented as part of a face composite-building system, appear to work well.

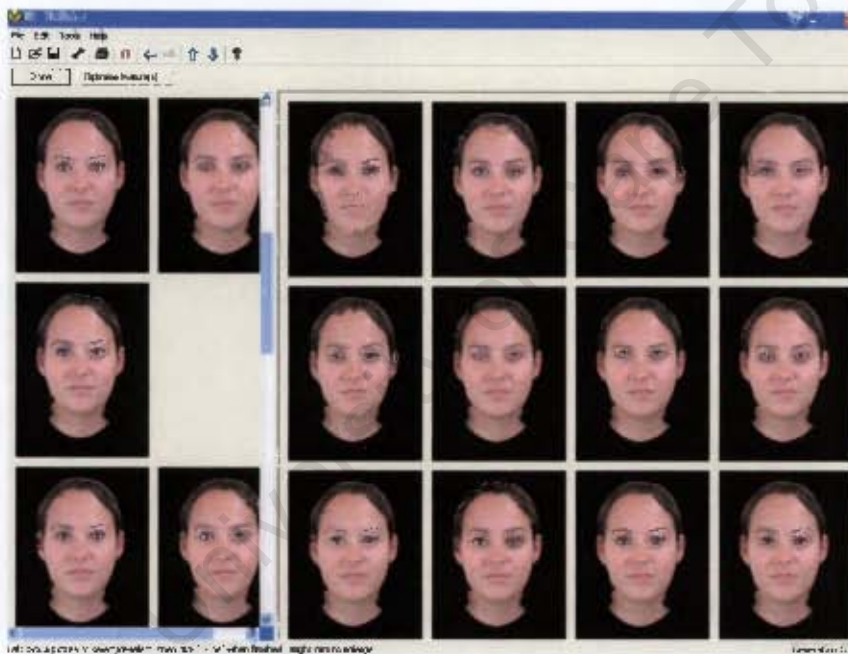
Although computerised systems are more sophisticated than the manual ones, it seems that a common obstacle for *all* composite systems at this time is they cannot produce good quality likenesses when reconstructions are done from memory (a forensically realistic condition).

Despite the disappointing empirical findings, composites are nevertheless invaluable to police inquests. Even with their shortcomings, they are useful for generating possible suspects, eliminating unlikely suspects in mug files or lineups prior to exposure to witnesses, narrowing the potential group of suspects and improving chances of identifying and apprehending offenders (Brignull, 1998; Davies & Valentine, 2007; Green & Geiselman, 1989; Prag, 2005). As a screening tool, these imperfect composites may nevertheless be accurate enough to alert a police officer to stop and question a potential suspect or avoid disturbing an obviously innocent individual (Koehn, Fisher & Cutler, 1999).

New computerised system: ID

Our superior ability to process faces holistically is at the heart of a new composite system called ID under development at the University of Cape Town. Given the problems associated with producing accurate composites using the aforementioned facial reconstruction systems, this study aims to test a new eigenface-based composite system called "ID" (Rosenthal, de Jager and Greene, 1998; Tredoux et al., 1999; Tredoux, 2001). This system is recognition-based, where the user chooses from a display of facial composites (called the graphical user interface or GUI, see Figure 1) rather than constructing their own.

Figure 1: The main ID graphical user interface (GUI) during a reconstruction



This system was developed to compensate for the numerous limitations inherent in current facial composite systems. It presents the face as a whole rather than fragmented features, which is consistent with the way people encode faces (Ellis et al., 1978a; Laughery & Fowler, 1980; Wells & Hryciw, 1984). It also removes the need for verbal descriptions – selecting in this way is much like picking criminals from a mugshot album, which avoids the task of having to describe the target face – and expert operators. Brace et al. (2000)

found about a 10% increase in naming rate when composites were created by an operator alone compared with composites constructed by the normal interaction process involving another person i.e. the witness. ID compensates for this by allowing the witness to operate the program themselves, independent of an operator. ID is capable of efficiently searching an extremely large sample space of alternative faces, and finding an accurate likeness in a relatively short period of time.

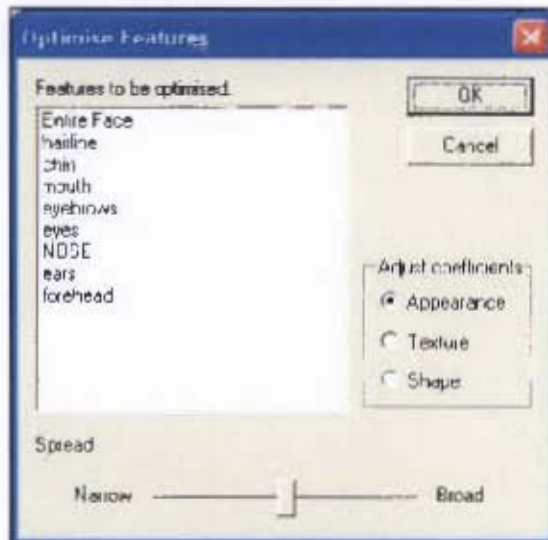
ID works on the basis of principal component analysis (PCA), a statistical technique that can extract the principal components or “eigenfaces” from a sample of faces and project these eigenfaces into the “face-space”, where the differences between face images are characterised and then compared to known faces for recognition (Turk & Pentland, 1991). When eigenfaces are recombined randomly, a face-like image that is not one of those in the original set may be accurately generated. The Euclidean distance between any 2 faces in the face-space can be used as an index of their physical and perceptual similarity. Reconstructions are produced by linear combination of eigenfaces, using coordinates as weights (Tredoux et al., 2006). This technique works well for faces, and is particularly valuable as it provides a set of photographic-like reference faces (eigenfaces) that can be holistically combined to produce a novel face, a vital underlying factor of any face evolution system (Frowd, Bruce, McIntyre, et al., 2006; Hancock & Frowd, 2001; Tredoux et al., 1999). As Hancock, Bruce and Burton (1998, p. 2278) state, “facial images are represented in terms of the statistical regularities within a set.” For a more detailed explanation of this computational basis, refer to Kirby and Sirovich (1990), and Turk and Pentland (1991).

Two optimisation techniques can be implemented when using ID, namely Population Based Incremental Learning (PBIL) or M-Choice. With PBIL, the witness selects one of the 12 faces displayed on the GUI that looks most similar to that of the target. A new set of 12 faces is then generated that looks more similar to the previously selected face, whilst ensuring a wide spread of

different-looking faces. The process is repeated in this manner until the witness is satisfied that they cannot obtain a better likeness of the target, whereupon the process is stopped and the face saved. With the more recently developed M-Choice algorithm (devised by Tredoux et al., 2006), the witness is not restricted to selecting just one face from the GUI that resembles the target's face, but can select multiple faces concurrently in each generation (or evolutionary cycle) (Prag, 2005; Tredoux, 2001). Users are able to go backwards and forwards during the reconstruction process, and can also resample thereby producing a new set of faces on the GUI. The resample button assists users by allowing them to obtain a new set of faces with which to work when they cannot find good enough likenesses in the current selection of faces. The user's generation history is shown on the left-hand side of the screen throughout the process so that the user can refer back to previous generations if they so choose. These can also be selected in combination with faces from the GUI to produce new sets of faces. The generation history can also be saved for future reference.

Witness feedback and direct observation indicated the need for the addition of some sort of featural search into the system to enable users to manipulate and refine the finally chosen eigenface. Therefore the "optimise features" tool was incorporated to the configural basis of the system, and is indicative of the dual processing strategy (see Figure 2). As Frowd (2001, p.27) asserts, "...since an exclusive holistic bias may not be the best system for analytical encoders, a hybrid holistic-componential photofitting approach may be optimal for a witness".

Figure 2: The optimise feature(s) tool



The optimise features tool allows individual features of a chosen face to be manipulated. For example, if the eyes of the finally chosen face are "not right" according to the witness, the optimise features tool can expand the range of eyes while keeping the rest of the face constant (see Figure 3). ID would produce a new generation of faces and the eyes on each of these faces would vary, while the rest of the features on each of these faces would remain exactly the same. The witness would then select the face with the most similar-looking eyes to that of the target. The tool can be activated and deactivated at will and the procedure can be administered to any of the facial features (Tredoux et al., 2006).

Figure 3: Effects of the optimise features tool on the eyes



when the target is well known and well memorised, but not when the target is unfamiliar and recently committed to memory.

It was noted that the failure of the reconstructions were not due to encoding or contamination effects as all participants correctly identified both targets (familiar and unfamiliar) from lineups before and after the reconstruction task. Therefore a third experiment was conducted that repeated the procedure of experiment 2 with two unfamiliar target faces. Six participants used E-Face to reconstruct two unfamiliar faces from memory. Two hundred and thirty judges evaluated the resulting composites. An overall identification rate of 43% was found, which is somewhat higher than the average rates recorded in other studies that have tested facial reconstruction from memory with composite systems. For example, Christie and Ellis (1981) found a 23% identification rate, while Ellis, Shepherd and Davies (1978) found only 13% accuracy. These favourable results reveal the promise of this eigenface-based composite system (Tredoux et al., 1999).

Following much experimentation with E-Face over a number of years, numerous problems were identified and solutions were subsequently developed, leading to multiple revisions to the software until a working prototype was produced and the system was renamed "ID". For example, a much larger database of higher quality was built, the ghost-like composites resulting from PBIL's scaling and centering solution were mapped to the average of the collection of faces resulting in a "shape-free face", the greyscale images were modified to colour images, and the MChoice algorithm was devised to allow multiple choices from the face array (as opposed to a single choice when using PBIL). Most recent improvements include the addition of the optimise features and accelerator tools. For a more detailed discussion of the many system changes made to ID, please refer to Tredoux (2001).

Previous studies have only assessed the white male database, thus as part of his honours thesis, Prag (2000) implemented the black and coloured male databases. Participants reconstructed 2 faces using the ID system either with

a photograph in view, or from memory after a two-minute videotaped exposure. Fifteen independent judges rated the resulting composites for similarity to their respective target and three foils, using a 7-point Likert scale. The average similarity rating score for the in-view reconstructions were 46% and 42% for reconstructions from memory. An average rating score of 50% was found for the coloured target and 37% for the black target. The average identification rates were 40% for the black target and only 11% for the coloured target. In a second study by Tredoux and Nunez, the black, white and coloured male databases were implemented. Average identification rates were 38%, 52% and 76% for the black, coloured and white databases respectively (Tredoux, 2001).

In 2005, Prag again investigated whether ID could produce satisfactory likenesses of faces from different but related face populations (Prag, 2005). Participants made in-view reconstructions of a coloured and Indian face, using one of three databases (coloured, Indian or coloured/Indian). Independent judges evaluated the quality of the resulting composites through a rating task. All reconstructions received an average similarity rating of 47% (which does not compare favourably with the 72% found by da Costa (1998) for in-view reconstructions of white male targets). However, the overall identification rate was 51%, which matches the findings of da Costa (1998).

On the basis of the results of the aforementioned studies and witness feedback, it was found that the decision to map face images to a common shape was very limiting; therefore shape was re-incorporated as a searchable and modifiable aspect. ID was revised to work with parallel shape and texture models. Now when a selection is made, both the texture and shape coefficients describing that face are read, and used to conduct parallel searches of texture and shape space (Tredoux, 2001). To ascertain whether this parallel shape and texture model had improved reconstruction capabilities, another experiment was performed. The enhanced version (texture + shape) showed a 50% identification rate for novice users (compared to just 22% using the original version) and an impressive 86% for expert users (Tredoux, 2001).

Studies suggest that ID can assist in the production of recognisable face composites from memory. However, it had not been investigated whether ID produces reconstructions that are more recognisable than featural composite systems currently employed by police worldwide. Prag (2005) made such a comparison, where the performance of ID was compared to a face composite system currently used by international law enforcement agencies, namely FACES. To test the optimality of both systems, expert users produced reconstructions in-view of white male target faces. These reconstructions were evaluated against novice in-view reconstructions by means of rating, sorting and ranking tasks. The FACES system performed better than ID for both the expert and novice in-view conditions in all three tasks, with FACES composites receiving a hit rate of 69% (compared to 54% for ID). In fact, the best rated expert composites of all four targets were created using the FACES system, as well as the best rated novice composites for 3 of the 4 targets (Prag, 2005). It seems that FACES has a clear advantage when composites are constructed in-view, however consider that featural manipulators are particularly well suited to reconstructions made in full view, as specific features can be compared and matched to those on the visible target.

To test the two systems under more realistic conditions, reconstructions were also made by novice users from memory, following a one-minute exposure to one of four targets during a live staged event. The resulting composites were evaluated against the in-view composites built earlier by rating, sorting and ranking tasks. An average hit rate of 42% was found. In general, the FACES composites were slightly better than the ID ones, receiving a hit rate of 45% (compared to 39% for ID). FACES performed better when the composites were built in view, but ID composites proved to be better for the two from-memory conditions (in all 3 tasks). The general conclusion was that FACES is better able to reconstruct faces in view, while ID is better able to reconstruct faces from memory, which is the more forensically valid condition and hence shows ID's promising potential in the field (Prag, 2005).

For my Honours research in 2004, I aimed to test whether guided memory techniques would assist memory of a target face such that the resulting face composite produced using the ID composite system would be an improvement on composites produced from memory alone. The guided memory interview was based on previous studies (which will be elaborated upon in the following section) and adapted to suit the particular situation of the study. Participants took part in a live staged event where they witnessed a white male for approximately 2 minutes and were required to return one week later to reconstruct the target face either from memory after a South African Police (SAP) interview, from memory after a guided memory interview, or with a photograph in view. Results were puzzling as there was no consistency among the evaluative tasks (See Table 1 below).

Table 1: Percentage of composites correctly matched to the corresponding target in a 6-person target present lineup

Target	In view	Guided memory	SAPS
1	69%	30%	22%
2	18%	18%	34%
3	13%	24%	39%

Similarity ratings of the target were also significantly associated to composite condition. Target 1 followed the hypothesised pattern, with in-view composites obtaining the highest ratings, followed by the guided memory composites and lastly the SAP composites. However, target 3 obtained unusual results, with the SAP composites obtaining the highest ratings, followed by the in-view and guided memory composites. This is in contrast to research findings that in-view composites are better than those produced from memory (Wolgater & Marwitz, 1991). The overall results were disappointing and showed that only 40% of participants had encoded the face in memory sufficiently to recognise it only one week later. I concluded that the poor results were due to poor encoding at the time of the event which subsequently led to the participants not being able to recognise the target from a lineup. For the current study, exposure duration was increased to approximately 20 minutes as previously the exposure duration may not have

The “accelerator tool” allows for the search gradient to be controlled. If a witness finds a face that looks quite similar to that of the perpetrator, the accelerator can be used to produce other faces similar to that particular face. It can also be used to produce a less similar spread of eigenfaces in a certain generation if the witness feels that they are getting too similar. This particular new feature aims to resolve a complaint lodged by several participants that the eigenfaces of the ID system become too similar too quickly when progressing from one generation to the next.

At present, the ID system comprises six databases¹, namely: (1) ‘black’ (African) male, (2) white male, (3) ‘coloured’ (mixed race) male, (4) Indian male, (5) white female and (6) ‘black’ female. Further databases are in the process of being built and will be available for testing in the near future. The male databases were the first to be developed and have been examined in past studies (Da Costa, 1998; Prag, 2000; Prag, 2005; Sullivan, 2004; Tredoux, 2001; Tredoux et al., 2006), but the female database has only recently been developed. Therefore this research assesses it for the first time.

Evaluation of the ID system

Work first started on this composite system in 1998. A series of three experiments² conducted by Tredoux et al. (1999) tested the first version of the ID system (previously called E-Face). In experiment 1, 15 participants used E-Face to reconstruct three faces with a photograph in view. 267 judges evaluated the resulting composites through matching and rating tasks. An overall identification rate of 51% was found, which clearly shows that the system is able to produce recognisable composites while the target is in-view. Experiment 2 aimed to evaluate the ability of E-Face to produce recognisable composites from memory with varying ‘memory strength’ – a familiar face of a famous actor and an unfamiliar unknown face. The identification rate for the familiar face was 58%, and 17.2% for the unfamiliar face. These results suggest that E-Face is able to produce recognisable composites from memory

¹ Participants were tentatively categorised by the race to which they identified themselves.

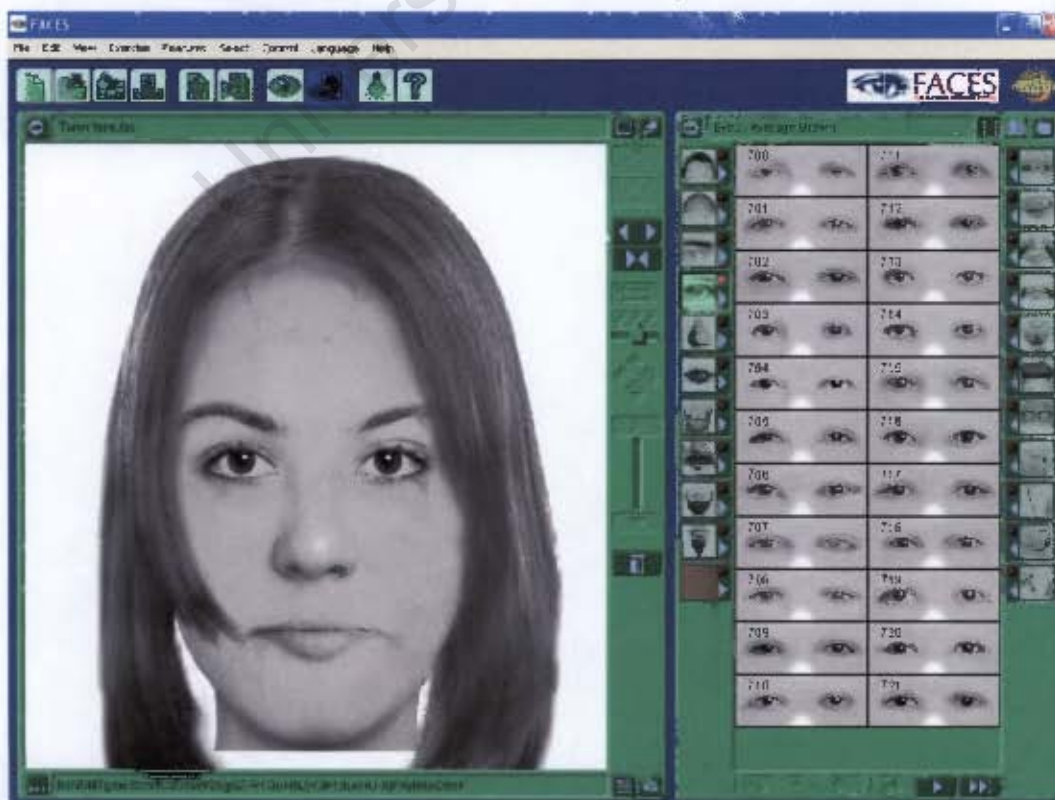
² Experiments 1 and 2 are detailed further in da Costa’s honours thesis (1998).

been long enough or the intervention not strong enough. This study hopes that the extended time period of witnessing the targets will be sufficient to encode the face adequately into memory for the 'witnesses' to reconstruct it after a considerable delay.

The comparative system in this study: FACES 4.0

FACES is a user-friendly system with a database of almost 4,000 specially coded black-and-white facial features contained in 22 feature catalogues: hair, head shapes, eyebrows, eyes, noses, lips, jaw shapes, moustaches, beards, goatees, skin tones, forehead lines, eye lines, smile lines, mouth lines, chin lines, head wears, glasses, moles, scars, piercings and tattoos. Adult faces of any race or gender between the ages of 17 and 60 can be created by automatically blending together all of the individually selected features, resulting in a composite resembling a good-quality black-and-white photograph (Prag, 2005).

Figure 4: The main FACES 4.0 user interface during a reconstruction



The selection panel on the right hand side of the screen contains miniature thumbnails which contain each of the 22 feature catalogues (see Figure 4). When a user clicks on one of the thumbnails, for example the lips icon, pages of various lips are displayed for the user to browse. A useful feature is the subcategory selection menu that appears if the arrow next to the feature icon is clicked. This menu lists different feature subcategories, for example thin, even or heart-shaped lips. The user can then narrow the search for the desired feature by choosing the relevant subcategory. Another handy feature of FACES is the previous/next selection arrows that allow users to view all previous feature selections.

The control panel displays various feature editing tools, such as a position slider that can move a feature up or down, an enlarge and shrink feature button, a 'flip' hair parting button and a hair shade slider which allows three levels of hair colour (light, medium and dark). There is also a 'delete selected feature' bin to discard a selected feature. Underneath the main display screen there is an ID profile button where information regarding the crime can be stored, such as the victim's personal details, incident number, police department, criminal(s) details, clothing and weapons. There is also a feature coding button that stores each feature code, its position and size for possible future reference.

Support for the effectiveness of the FACES system is found in Prag (2005), who found that the majority of the highest performing composites in all the respective conditions of his study were produced using FACES. He concluded that FACES was better at reconstructing faces in view, while ID was better at reconstructing faces from memory. Unfortunately there are very few empirical studies where the effectiveness of FACES has been tested.

Frowd, Carson, Ness, McQuiston-Surrett, et al. (2005) compared composites from several systems in use worldwide under realistic conditions, specifically PROfit, E-fit, artist sketches and FACES, as well as EvoFIT. Participants constructed a composite using one of the 5 systems 2 days after viewing a photograph of a celebrity. The construction procedure closely matched that

found in police work as experienced composite operators and a Cognitive Interview were utilized. Composite naming was surprisingly low (only 3.3% correct overall), with sketches named best at 8.1%, followed by EvoFIT (3.6%), FACES (3.2%) and PROfit (1.3%), with no E-fit composites being named. A similar conclusion was reached by Koehn and Fisher (1996) using Mac-A-Mug Pro. In addition, composites were sorted overall to an accuracy of 42%. Sketches performed best once more (54.4%), followed by the other systems at about 40%. Frowd et al. concluded that E-fit, PROfit and FACES are very similar systems and were found to perform equivalently.

In the current work of Frowd et al. (in press), two popular computerised systems in the US, FACES and Identikit 2000, were evaluated against PROfit. In Experiment 1, witnesses constructed a composite with both PRO-fit and FACES using a realistic 2-day delay between composite construction and recognition. The resulting composites were very poorly named (overall correct naming was 0.5%), but the PRO-fit composites elicited significantly more correct names when cued than those from FACES (9.2% versus 2.5%). The sorting task showed 31.4% overall correct and was also significantly higher for PRO-fit (37.5%) than for FACES (29.2%). Overall mean likeness ratings once again favoured PRO-fit composites (3.3 out of 5) than FACES composites (2.5 out of 5).

With the FACES system, witnesses select 'isolated' features out of the context of a whole face and these features are contained in mixed race and gender feature libraries (Frowd et al., 2005). Hence, it was predicted that participants would be exposed to irrelevant features (e.g. male hairstyles when female styles were required), which would lead to confusion and ultimately poor composites.

5. Context reinstatement

There are situations daily where people have to remember a face of someone, perhaps the mugger who stole their purse or a burglar who robbed them at gunpoint. Therefore eyewitness identification is of crucial importance to everyday life. For example, it is imperative in a court of law for a witness to be able to recognise a face as familiar and to recall where this face was seen, as these identifications are important to law enforcement officers for capturing criminals, and also to jury members for providing evidence on whether a suspect is guilty or not (Brigham, Maas, Snyder & Spaulding, 1982; Gibling & Davies, 1988).

It is well documented in memory research that human recognition performance is imprecise and easily influenced by situational constraints.

“Eyewitness identification is of uncertain accuracy at best and profoundly wrong and harmful at worst”

(Malpass & Devine, 1981, p.343)

Therefore it's important to identify procedures that might improve the reliability of eyewitness identifications and to find ways of strengthening techniques for obtaining and evaluating eyewitness testimony in order to enhance accuracy and reduce errors. One promising approach involves procedures designed to reinstate the context surrounding an event (Cutler et al., 1987; Malpass & Devine, 1981).

In a strict sense, all remembering – both recall and recognition – is cued. Memory for a past event never materializes from nowhere, but rather takes place within a specific context and is influenced by the information contained within a stimulus. If contextual features are encoded with the stimulus to be remembered, they can act as potential aids to memory at the time of retrieval (Krafka & Penrod, 1985).

When recognition is requested after lengthy periods of time, the accuracy of the witnesses' recognition can be enhanced by reinstating the context of the

witnessed event (Malpass & Devine, 1981). Context may provide retrieval cues for memory as long as the contextual information has been encoded and is also most effective when memory for the to-be-identified person is impaired (Cutler & Penrod, 1988; Krafka & Penrod, 1985). If important cues are degraded due to impoverished encoding (such as disguise), storage factors (such as a long retention interval), or retrieval conditions (such as suggestive instructions), then contextually reinstated cues may improve the accuracy of identifications by enhancing the witness's ability to recognise the target (Cutler et al., 1987).

Several dimensions of 'context' have been identified (Krafka & Penrod, 1985):

- Physical environment – physical features of the location become associated in memory with the studied material and these physical features later serve as contextual retrieval cues to memory.
- A person's emotional frame of mind at the time of encoding – there is support that moods may cue remembered material of a similar disposition.
- Environmental context (Smith, 1979) – refers to the physical surroundings in which an event occurs, including location, objects and people present, odours, sounds, temperature and lighting.

Environmental, emotional, and other contextual information is encoded into memory, together with the target face, and this encoded material becomes associated in memory with features of the setting. These features later serve as contextual cues (Cutler et al., 1987).

Guided Memory techniques

A guided memory interview is similar to and derived from the Cognitive Interview, but focuses more on the person's psychological (cognitive and emotional) state at the time of the event. It takes the witness step by step through the events leading up to the offence and the event itself. This procedure reminds witnesses of their reactions to the event, their thoughts

and feelings of the perpetrator, the scene and its surroundings (Cutler & Penrod, 1988; Gibling & Davies, 1988; Krafka & Penrod, 1985; Malpass & Devine, 1981). It is not an in-depth probing of a person's memory of a face, but is an indirect method of extracting the memories without the constraints of closed-ended questions.

Studies have shown positive results regarding reinstatement of context, and that guided memory techniques led to greater identification accuracy and had a significant effect on memory and recognition performance. This line of research can be traced back to a series of experiments conducted by Smith and colleagues on verbal learning. Smith (1979) believed that memory is best when the situational or contextual conditions present at learning are reinstated at the time of the test. He showed a very strong enhancing effect on recall by returning subjects to the original learning environment (where they learned word lists) for testing, which he aptly termed the 'environmental reinstatement effect'. Furthermore, he found that subjects who were asked to return mentally to the study environment were as effective as those who physically returned. Smith (1984) found that subjects instructed to recall the learning context recalled 21% more words than the uninstructed subjects, consistent with Smith (1979). Winograd and Rivers-Bulkeley (1977) also found that changing context from study to test significantly impaired recognition.

As a result of Smith's findings, Malpass & Devine (1981) adapted this procedure to the process of obtaining eyewitness identifications after long delays. They examined the effect of the guided memory on rates of choosing and accuracy. Subjects witnessed a live, staged event of vandalism and were contacted five months later to identify the vandal from a photographic lineup. Those subjects given guided memory interviews were reminded of the events of the evening of the vandalism, and their feelings, memory of details of the room, memory of the vandal and their reactions to the events were explored. Context reinstatement through the guided memory procedure increased the rate of accuracy from 40% to 60% correct identifications after a 5-month delay. This increase is a sizeable improvement in recognition accuracy.

Based on Malpass and Devine's findings, Krafka and Penrod (1985) conducted their own guided memory field experiment to test reinstatement of contextual information on eyewitness identification performance. Naïve store clerks were asked to identify a customer encountered either 2 hours or 24 hours previously from a collection of photographs. The event was designed to be an infrequent but not improbable event (the 'customer' had paid for a small item by traveller's cheque), so as to increase the likelihood of remembering it. The procedure used to reinstate context nearly doubled accurate identifications from 29,2% to 55% correct after a 2hr delay, and increased from 39,1% to 42,9% after a 24hr delay, proving that guided memory has a significant effect on memory and recognition performance. Cutler et al. (1987) also examined the reinstatement of context relative to eyewitness identification accuracy and found that reinstatement of environmental and emotional state context enhanced identification performance, in line with the results of Krafka and Penrod (1985) and Malpass and Devine (1981).

Davies and Milne (1985) also demonstrated that subjects given a guided memory interview produced composites that led to more accurate information than those who produced composites from a verbal description alone. 32 housewives were given different physical reinstatement (same room vs. different room at testing) or mental reinstatement (spontaneous recall vs. guided memory of the original event) to assess the effects on memory for an intruder. One week after interacting with an intruder, the women were asked to provide information regarding the intruder's appearance. Half the subjects were interviewed in the same room as the event, while the other half was interviewed in a new and different room. In addition, half the subjects were given a guided memory interview, prior to reconstruction. Davies and Milne replicated the environmental reinstatement effect mentioned by Smith (1979) as irrespective of testing room, guided memory increased the quality of composites produced by the witnesses.

In a study by Gibling and Davies (1988), subjects' memory for details of the events leading up to the incident (pre-event information), and then memory

for the incident itself was probed. Subjects were asked to recall step-by-step the sequence of actions performed by the target and by any other people present. To aid recall, environmental context was reinstated by means of slides of the location. Overall accuracy of identification increased from 32% to 49%, showing a similar increase to the 20% reported by Malpass and Devine (1981).

Context reinstatement has also been found to reduce the effect of misleading post-event information. Bekerian and Bowers (1983) demonstrated that reinstating the appropriate contextual cues eliminated the influence of misleading post-event information upon memory. Gibling and Davies (1988) examined the influence of misleading face composites upon eyewitness memory. Subjects viewed a videotaped incident and subsequently received a composite of the target designed to depict either correct or misleading information regarding his appearance. They found that although the context reinstatement did not entirely eliminate recall of the misleading features, it did significantly reduce the effect from 73% to 60%. Meissner and Brigham (2001a) reiterated this sentiment, who found a significant improvement in recognition accuracy following composite reconstructions across 15 studies.

My Honours study aimed to test whether guided memory techniques would assist memory of a target face such that the resulting face composite produced using the ID system would be an improvement on composites produced from memory alone. The guided memory interview was based on previous studies. Physical, environmental and emotional cues were adapted from previously documented guided memory interviews to suit the particular situation of this study. The participant was reminded of the events surrounding the meeting of the man. The participant's memory for the details of the events leading up to the meeting was explored. Environmental context information was included in the interview as Smith (1979) found that it was a source of useful retrieval cues to aid in recalling information learnt in a specific context. They were asked to silently recall details such as how they felt that day prior to the meeting, where in the room they were sitting relative to the other participants, the room's characteristics and so on, and then

asked to recall the meeting itself (Gibling and Davies, 1988). No explicit mention was made of the target's appearance throughout questioning, however questions regarding the context of his appearance were asked to evoke memories of the perpetrator's appearance without actually stating any details about his appearance. Participants were asked not to answer the questions but to silently imagine and visualise the event in their minds. The results were disappointing: Target 1 followed the expected and predicted pattern, with in-view composites obtaining the highest similarity ratings, followed by the guided memory composites and lastly the SAP composites. Target 3, however, showed an unexpected pattern, with SAP composites obtaining higher ratings than the in-view condition. This should not be the case, as composites constructed with a photograph in view should have gained higher scores than composites produced from memory. I concluded that the poor results were due to poor encoding at the time of the event (only 40% of participants had encoded the face in memory sufficiently to recognise it only 1 week later) which subsequently led to the participants not being able to recognise the perpetrator from a lineup. The guided memory techniques are a fairly new and under-researched area of research and it is difficult to make assumptions regarding results of this study as one does not know if the interview contained too many or too few contextual cues, or if the length of the interview had any effect. The event may have not been encoded properly in the first place, therefore resulting in poor composites as participants had no memory of the face.

In the current study, the event is of a con-artist situation where participants will view the target for about 20 minutes. The exposure duration may not have been long enough or the intervention not strong enough in my previous study, therefore this study hopes that the time period of witnessing the targets (20 minutes) will be sufficient to encode the face adequately into memory for the 'witnesses' to reconstruct it 5 weeks later.

The Cognitive Interview

Geiselman, Fisher, MacKinnon and Holland (1986) recognized that the standard interview protocol used by police during an investigation consists of specific, closed-ended questions with no retrieval mnemonics being utilised. They subsequently developed the Cognitive Interview as a memory retrieval procedure to assist eyewitness recall of events. It is based on a number of memory-retrieval mnemonics designed to facilitate the recall of as much unbiased information as possible regarding a crime, including a guided memory interview (see points 1 and 2), and incorporates nonleading, open-ended questions with a combination of four retrieval techniques:

- 1) Context reinstatement
- 2) Emphasising to the witness the importance of reporting everything they can remember even if the information seems unimportant
- 3) Recalling the events in a variety of orders
- 4) Mentally changing perspectives.

Context reinstatement has been shown to improve eyewitness identification accuracy. Asking the witness to report every detail may lead to important facts being remembered that occurred along with unimportant details. Recalling the events in order may lead to some people reconstructing what they *think* happened based on knowledge of similar cases. The act of recalling the events backwards forces the witness to actively reflect on the event therefore reducing inaccurate or incomplete reports (Geiselman et al., 1986). Mentally changing perspectives may also enhance the completeness of the reports.

The Cognitive Interview was an improvement over standard interviewing techniques, but it did not take into account some of the practical obstacles encountered by police in the field. For example, real victims of crime are often emotional and anxious, and have poor communication skills. The original Cognitive Interview was then revised to address issues such as controlling the emotional state of the witness and improving communication between the witness and police investigator (Fisher, McCauley & Geiselman, 1994). The

revised Cognitive Interview expands on the four memory-retrieval techniques of the original interview, and incorporates techniques such as encouraging the witness to use mental imagery, the interviewer tailoring the questions to suit the particular witness (witness-compatible questioning), and encouraging witnesses to use non-verbal responding to supplement their verbal responses (eg. Drawing the location or acting out a movement).

Most research has shown the original and revised Cognitive Interviews to be very effective at increasing the amount of correct information recalled by witnesses. Geiselman and his colleagues (1986) reported a 17% increase in recall following the Cognitive Interview, a significant enhancement to eyewitness memory, and believed that the beneficial effects of the Cognitive Interview lie in the guided memory components of the interview which encourage eyewitnesses to mentally reinstate the contextual elements present at the time of the original event. In a field test of the Cognitive Interview, Fisher, Geiselman and Amador (1989) compared seven experienced police detectives who were trained to use the Cognitive Interview technique with nine untrained detectives. They found that the trained detectives obtained 63% more correct information than did the untrained detectives. This field study replicated the results of previous laboratory studies and demonstrated its practical utility in real-world settings.

Bekerian and Dennett (1993) conducted a meta-analysis of 12 major studies in the area of Cognitive Interview techniques. This revealed that the amount of correct information recalled with the Cognitive Interview increased anywhere between 12% and 92%, relative to a standard interview.

The preceding studies have shown the usefulness of the Cognitive Interview in aiding eyewitness recall of *information* regarding an event, but its application to face composite production has also been assessed. Luu and Geiselman (1993) attempted to enhance face composites by using the Cognitive Interview with the Field Identification System (FIS). Results showed that the Cognitive Interview improved likeness rating when using holistic processing, where features were examined in the context of other features. This positive

effect, coupled with Davies and Milne's (1985) findings suggests that the Cognitive Interview may be successfully employed to improve face composites.

Koehn et al. (1999) were concerned about the potential verbal overshadowing effect that the interview may have on witnesses' face memory, since Schooler and Engstler-Schooler (1990) found that verbalisation about a stimulus can impair later recognition. These concerns were however discounted by Meissner and Brigham (2001a) in their meta-analysis of verbal overshadowing in face identification. It revealed that the verbal overshadowing effects are quite small, and are specific only to certain experimental manipulations. Their analysis showed that overshadowing effects were more likely to occur when the identification task immediately followed the description task or shortly thereafter (less than 10 minutes), and when participants were given an elaborative instruction as opposed to a standard instruction during the description task. Finger and Pezdek (1999) also found that release from verbal overshadowing was possible by inserting a delay as short as 24 minutes between verbal description and face identification, which is far less than is typical in police work due to availability of police resources (Currently in South Africa, there is a minimum delay between criminal event and composite construction of 48 hours, but in almost all cases this delay is far longer). This led them to believe that the verbal description does not overwrite the visual memory of the face but instead makes it less accessible at the time of face identification. If the recognition task was done some time after the verbal description was generated, participants were less likely to rely on their verbal memory for the face than the original visual memory.

Koehn et al. (1999) nonetheless revised the Cognitive Interview to be used with the visual task of composite production, rather than the verbal task of description. They revised the interview in two ways, namely the promotion of pictorial processing and encouraging witnesses to think of any trait judgements they may have inadvertently made about the face. Kerr and Winograd (1982) tested the effects of elaborated verbal encoding contexts on face recognition. They found that the most improved recognition occurred

when the initial encoding context was reinstated at the time of recognition. They also note that memory for a face is enhanced when judgements are made about a suspect's character traits. Trait judgements and labels assigned at encoding are helpful at retrieval for recognising faces (Chance & Goldstein, 1976; McKelvie, 1976).

Recent work by Frowd and colleagues seems to validate the modifications made to the original Cognitive Interview to suit facial composite production mentioned above. Frowd, Bruce, et al. (in press) investigated this possibility of enhancing recognition through the use of a Holistic Interview, a procedure that required witnesses to make a series of personality judgments about the target face. Although composites were constructed immediately after exposure to the target face, they found better quality composites using the Holistic Interview than a Cognitive Interview.

A study by Frowd, McQuiston-Surret, Kirkland and Hancock (2004) extends this work by comparing the effectiveness of these two interviews over a longer retention interval. In the study, mock witnesses constructed a composite two days after viewing a target face using the PROfit system. They were either given a Cognitive Interview, a Holistic Interview or no interview prior to construction. They could not find a difference in composite quality over all conditions in a variety of evaluative tasks, so in a last attempt they removed the hair from the composites, as some participants had reported that several composites were identifiable from their hairstyles. The results of this re-run identification task revealed an advantage for the Holistic Interview, replicating the basic findings of Frowd, McQuiston-Surret, et al. (2004). These two studies suggest that the Holistic Interview can produce more identifiable composites irrespective of retention interval, relative to a Cognitive Interview.

The studies mentioned above have shown that the Cognitive Interview produces more details and more accurate information recalled compared to a police interview, and is effective in aiding face composite production. For context reinstatement to be effective however, it seems there must be an appreciable retention interval or impairment of memory (as is the case with

almost every real-life criminal case). Results suggest that improved retrieval through context reinstatement is not limited to laboratory research, but is a potentially valuable technique whose effects can be applied to real world situations with success (Gibling & Davies, 1988).

6. Multiple composites and morphing techniques

Not much literature on this topic has been generated, but those that have delved into the effects of multiple composites have shown positive results. Presenting multiple composites for recognition is known to elevate performance. Situations already exist where multiple composites of the same suspect have been created, for example, if the suspect has committed several crimes such as serial murder – the “Yorkshire Ripper”¹ being one infamous example. Although the composites were initially created as part of separate investigations, they can become linked once evidence identifies a single person as the perpetrator of the various crimes (Brace et al., 2006). Another example where multiple composites may exist is when there is more than one witness to a crime, which is a frequent occurrence in reality, such as a bank robbery.

Until recently, the police have been wary about allowing more than one composite to be released to the media since composite quality tends to vary considerably across witnesses and different attempts at the same target may look like quite different people. Hasel and Wells (2005, p. 1) cite an incident where three separate composites were released:

A convenience store was robbed and three eyewitnesses individually worked with police to produce facial composites of the culprit. Each produced a very different face, so the police released all three

¹ After his capture, seven composites produced during the investigation were compared to a photograph of the Ripper. It became apparent that four of the seven composites bore no resemblance whatsoever, and the remaining three were slightly similar to him. Only one composite bore reasonable resemblance to the Ripper and may have been useful during the investigation (Davies, Ellis & Shepherd, 1985).

composites on a wanted poster. Police later received a message from a nearby small town police department that read: "Have arrested two of the suspects and are hot on the trail of the third."

These issues can be avoided by broadcasting a morph, as it will tend to emphasise the facial characteristics on which the witnesses agree and thus most likely to be correct (Frowd, Bruce, Storås, Spick & Hancock, 2006). Morphing goes back over 100 years to Galton, who tried to create an average criminal face (Sergent, 1984; Young et al., 1987).

Bennett, Brace, Pike and Kemp (1999, in Bruce et al., 2002) reported enhanced identification when multiple composites were shown, compared to when a single composite was shown. Bruce, Ness, Hancock, Newman and Rarity (2002) explored whether combining composites (by morphing) of the memories of several individual witnesses would enhance composites' effectiveness further. Merging the different individual composites should tend to reinforce correct features of the face and deemphasise incorrect features. Participants constructed two composites, one with a photograph in view and another from memory, using the PROfit system. Before constructing the from-memory composites, participants were given a Cognitive Interview. Composites were morphed in pairs for each condition, and these '2-morphs' were in turn morphed together in pairs to produce a '4-morph'. The data showed that the original individual composites performed lowest, with the 2-morphs performing higher and the 4-morphs obtaining the highest similarity ratings, and this trend seemed stronger for the unfamiliar faces. The 4-morphs were rated as well as the best individual composite and significantly better than the average individual composite.

Brace, Pike, Kemp, Turner and Bennett (2006) also evaluated whether presentation of multiple face composites improves identification of a target face. Judges familiar with the target were presented with one, four or eight composites depicting the same target. The overall results showed that presenting more than one composite increased identification rates, and interestingly, the highest rates of identification were achieved by showing *four*

composites. A possible explanation for the superiority of showing four rather than eight composites is that a person may be able to make use of the similarities and differences of the multiple composites presented to them to extract additional information. However, if too many are shown, the variety is too large or there are too many images to compare simultaneously, and the task becomes too complex.

Recent research has suggested that a better representation of a target face can be obtained by constructing more than one composite of a face and then combining the individual attempts into a single morphed image. Frowd, Bruce, Storås, Spick and Hancock (2006) hoped to replicate the laboratory findings of Bruce et al. (2002) in a genuine criminal investigation of a series of sexual assaults in England, namely Operation Mallard. Three composites of the suspect had been constructed during the investigation, and although none of these led to the capture of the perpetrator, a conviction due to DNA evidence was successful. The individual composites from the case and the resultant morph of these were evaluated against a photograph of the convict. The results of four experiments suggested that the quality of the morphed image tended to be at least as good as the best individual composite or better, reproducing the general finding of Bruce et al. (2002), but have the advantage that the composites were constructed under more realistic conditions.

Hasel and Wells (2005) found that morphs were rated as more similar to the target face than the individual composites and as good as the best individual composite. However, the morph also showed a strong prototype effect, where morphed composites came to resemble non-target faces more than the individual composites. In addition, composite morphing produced an attractiveness bias – the morphed composites were considered more attractive than the mean attractiveness of the individual faces – and this advantage of the morph over the individual composites decreased as target attractiveness decreased. However, even when the prototype effect and the attractiveness bias were controlled for, a valid morph-superiority effect remained. Results once again support Bruce et al.'s (2002) belief that there is a morph superiority effect for composite faces.

Chapter 3

Methodology

Overview

The first part of the study consisted of three phases: witnessing an event, reconstruction phase and evaluation phase. Participants attended a staged event where they witnessed a female 'numerologist' for 20 minutes. Five weeks later they were asked to return to create a composite of the woman using one of two selected computerised composite systems, namely FACES or ID. Participants were randomly allocated to three conditions, namely in-view, from memory with a South African Police interview, or from memory with a guided memory interview. At the reconstruction phase, participants were shown how to use the assigned system, and were given time to practice reconstructing a trial face. Thereafter participants were instructed to reconstruct the perpetrator's face. The second part of the study tested whether morphing composites produced by multiple witnesses of a target led to greater identification than multiple composites used individually. Hence, the guided memory composites for each perpetrator were morphed, creating an additional six composites to be evaluated. The complete set of reconstructions was then evaluated using matching and rating tasks.

Stage 1: Database building

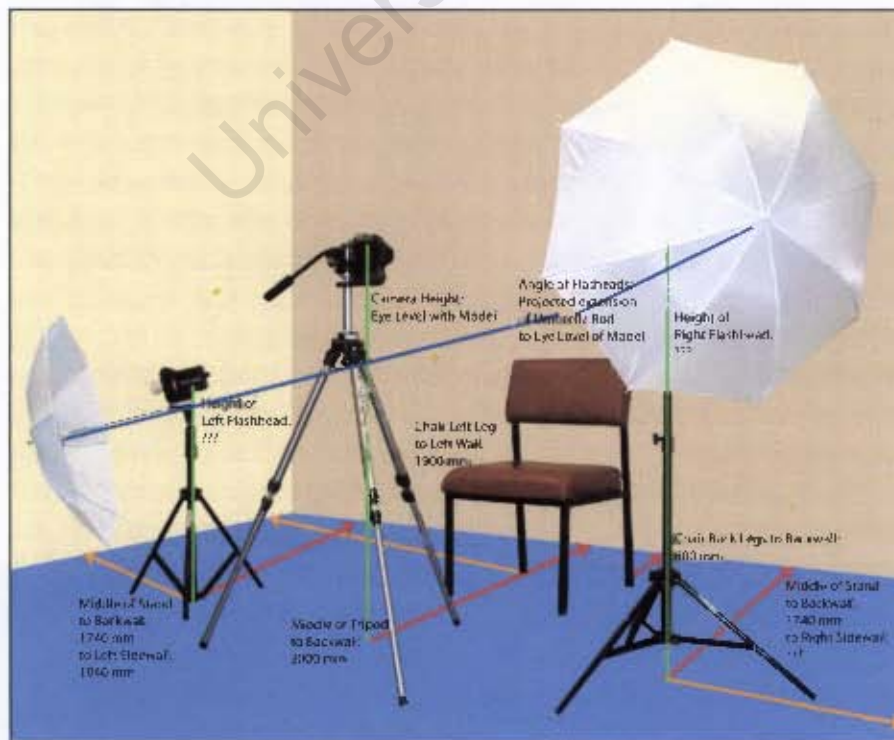
At the time of running this study, the ID system did not contain a white female database. Therefore to assess the system using a white female perpetrator, a new database needed to be created. This section illustrates the detailed process undertaken over several months.

The study was advertised around the university campus, using posters asking for female volunteers to have their photograph taken by the psychology department for a face database to be used for psychology experiments. About 250 white female faces without glasses, jewellery or makeup were carefully

photographed in a frontal, $\frac{3}{4}$ and profile face pose with a neutral expression, as well as a frontal smiling pose. The set of face images were standardized with respect to position, lighting and size. As PCA is very sensitive to changes in ambient lighting, we used a pair of flashlights (positioned at approximately 45 degrees and 1,428 meters from the subject; the camera was the same distance away; see Figure 5) and a small camera aperture ($f=8.0$). These photographs were taken at the University of Cape Town in an unused lecture room by several members of the ACSENT⁴ Laboratory. Although sampling was opportunistic (volunteer university students), we were able to collect a satisfactory number of faces.

A detailed protocol was developed by Schaupp and Tredoux, members of the Psychology Department at UCT, where equipment placement was precisely measured and recorded in order to ensure standardisation of every photograph taken. Specific camera setup details were noted, as well as lighting and camera lens angles and instructions for directing the subjects' facial positioning. A detailed version of this protocol can be viewed in Appendix A.

Figure 5: Graphical representation of the camera setup protocol used



⁴ Applied Cognitive Science and Experimental Neuropsychology Team

Subjects were first asked to remove any jewellery and/or glasses and to tie or pin their hair back so that it did not cover their ears or neck. Subjects were then asked to sit upright in the centre of the chair with their feet facing forwards towards the camera. The subject was informed that four photographs would be taken, and were asked to keep their mouths closed, display a neutral, non-smiling facial expression and look directly into the camera lens. While looking through the camera lens, the experimenter/photographer directed the subject to re-position their head in order to obtain symmetry as perfect as possible (see Figure 6). For example, vertical orientation of the subject's face was adjusted by instructions such as "please lift your chin slightly" or "lower your chin a little". Eye line was also checked to ensure that each pupil was horizontally in line, and if required to be adjusted, subjects were asked to tilt their heads either to the left or right until their pupils were in a straight line. Every instruction given to the subject was demonstrated by the experimenter/photographer so that the subject understood what was being asked of them and to reduce confusion and time wasting. This procedure was repeated for each photograph taken and the entire process took approximately 5 minutes per subject. The subjects involved were thanked for their participation and entered into a raffle draw to win an Apple iPod in exchange for their photographs.

222 of these faces were cropped from their original size of 3072 pixels (wide) x 2048 pixels (high) to 1822 pixels (wide) x 2376 pixels (high) and converted to 32 bit RGB colour images at a resolution of 180 pixels per inch. The remaining images could not be included in the database but were used elsewhere for testing purposes. Each facial image was loaded individually into Adobe Photoshop, where a face template was designed using paths to carefully and precisely place points (which make up the face template) on or around the various features of the face. Key facial landmarks in each face were located and adjusted to follow precisely the form of each feature, and then a new shape and texture PCA model was built using the 222 faces. Faces are represented well with PCA, but this is not the case for hair. Therefore all photographs were edited so that the hair was cropped as if it had been tied up in a ponytail. Those faces with excess hair which covered certain

features, for example, a fringe that covered the eyebrows (see Figure 7) or pieces of hair that had fallen across the ears, were excluded from the database as the resulting eigenfaces would be compromised. (Where possible the unwanted hair was removed using Adobe Photoshop tools). This is something that needs to be taken into account for further database building – to make sure that all excess hair is tied or pinned back so that it does not obscure any features.

Figure 6: An example of incorrect (a) and correct (b) facial alignment



Figure 7: An example of hair that could not be edited out (face excluded from the database)



Figure 8: An example of the face outline path being applied and altered to fit the face. The handles seen in (b) are used to control the curve of the path, allowing precision accuracy.

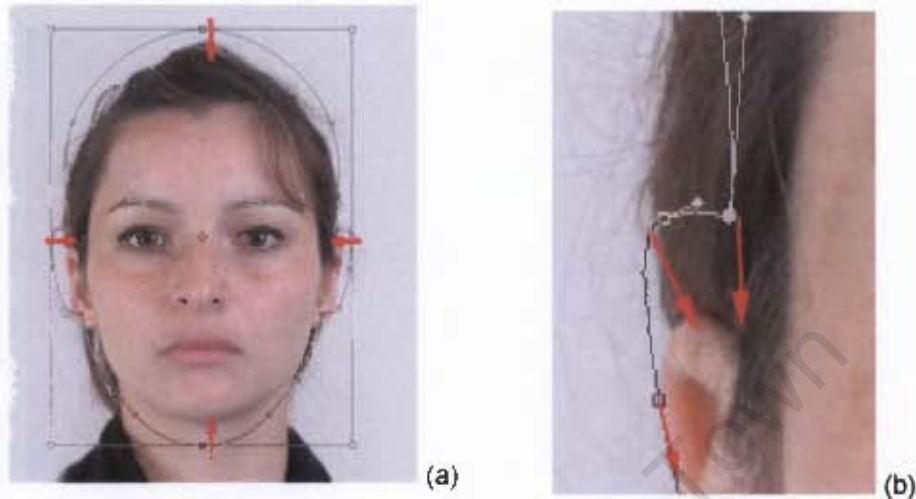


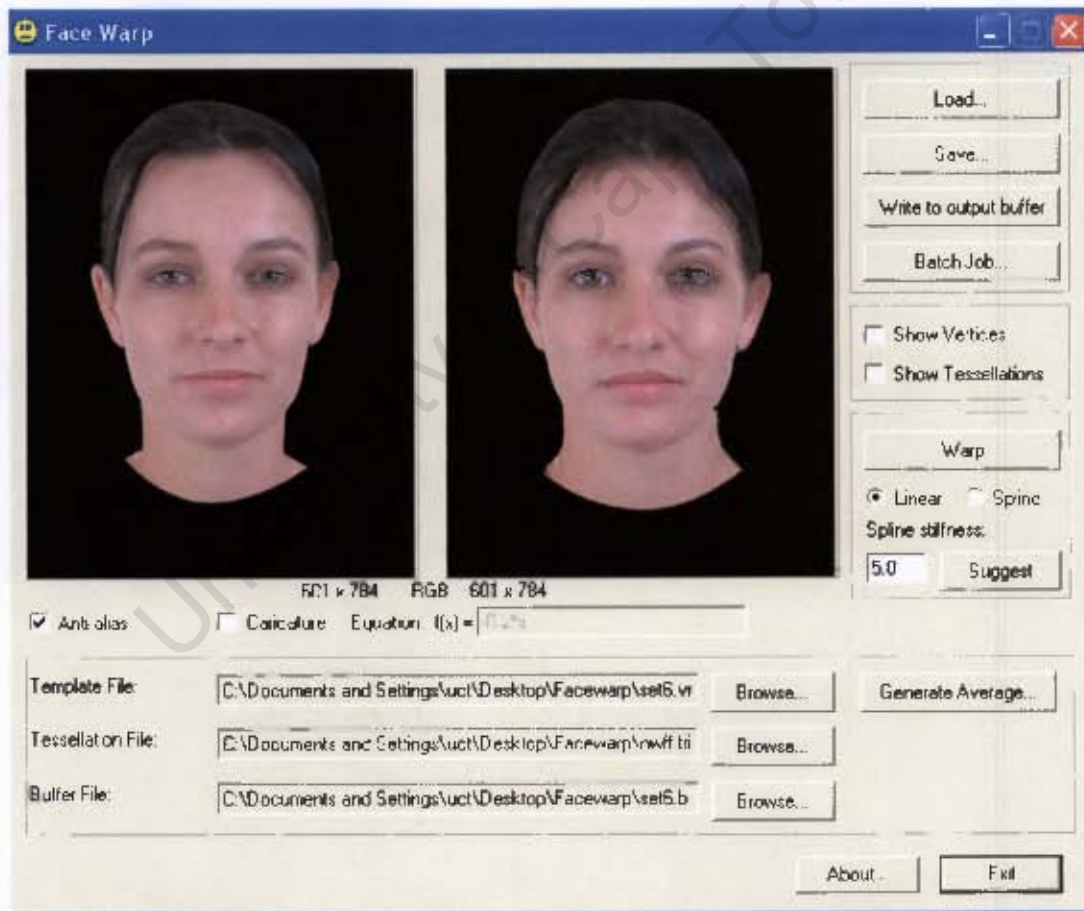
Figure 9: A face before and after landmarking. Note the hair has been cropped to a tied back style



Once the landmarking phase was complete, the images were scaled down to a more computationally manageable size of 601 x 784 pixels (72 pixels per inch) and warped using a specially developed computer-based programme called FaceWarp. Faces must be normalised in order for them to be processed as data points in PCA. Therefore, 'texture' and 'shape' elements of the face need to be extracted separately. The landmarked points define the

shape and also act as vertices for a set of triangular tessellations which make up the texture. These faces are then warped to the average face shape (Tredoux, Nunez, Oxtoby, Rosenthal, da Costa & Prag, 2001). FaceWarp was used to generate an average face for the 222 faces that made up the database (see Figure 10). Once the warping process was complete, PCA was conducted on the normalised faces, to derive "eigenfaces". Each eigenface was then saved to the database.

Figure 10: The main interface of the FaceWarp programme in use. The image to the left is the landmarked photograph (inputted image); the image to the right is the warped average of the inputted image



Stage 2: Event phase

Method

Participants

Seventy two white undergraduate students from the University of Cape Town (UCT) participated in witnessing the event and the reconstruction phase. Other race groups were not included in these phases to avoid race having any effect on the composites. It is well known that a cross-race effect exists in face recognition and therefore was avoided in this study (see Meissner & Brigham (2001b) for a detailed discussion). The participants had volunteered to take part by filling in their names and contact details on posters around campus. Participants were divided into three groups of 24 people for the event phase, with each group coming on separate days and witnessing a different perpetrator. Therefore three live events were staged using different perpetrators in each. The ages of the participants ranged from 18-25 years, with an average age of 20.5 years old ($SD=1.6$), with 34 male and 38 female participants. Participants were paid R50 for their effort and co-operation. Participants were not given specific details about their roles in the experiment, but were told prior to participating that the study would involve numerology. Participants were misled⁵ as to the real reason for research so as not to bias the results. The study aims to test the participants' memory of the target face after a significant delay, in an attempt to replicate a real crime situation where the perpetrator is seen for a reasonable amount of time. Therefore informing the participants beforehand of the intent of the study would lead to the participants paying more attention to the perpetrator's face and contaminating the results.

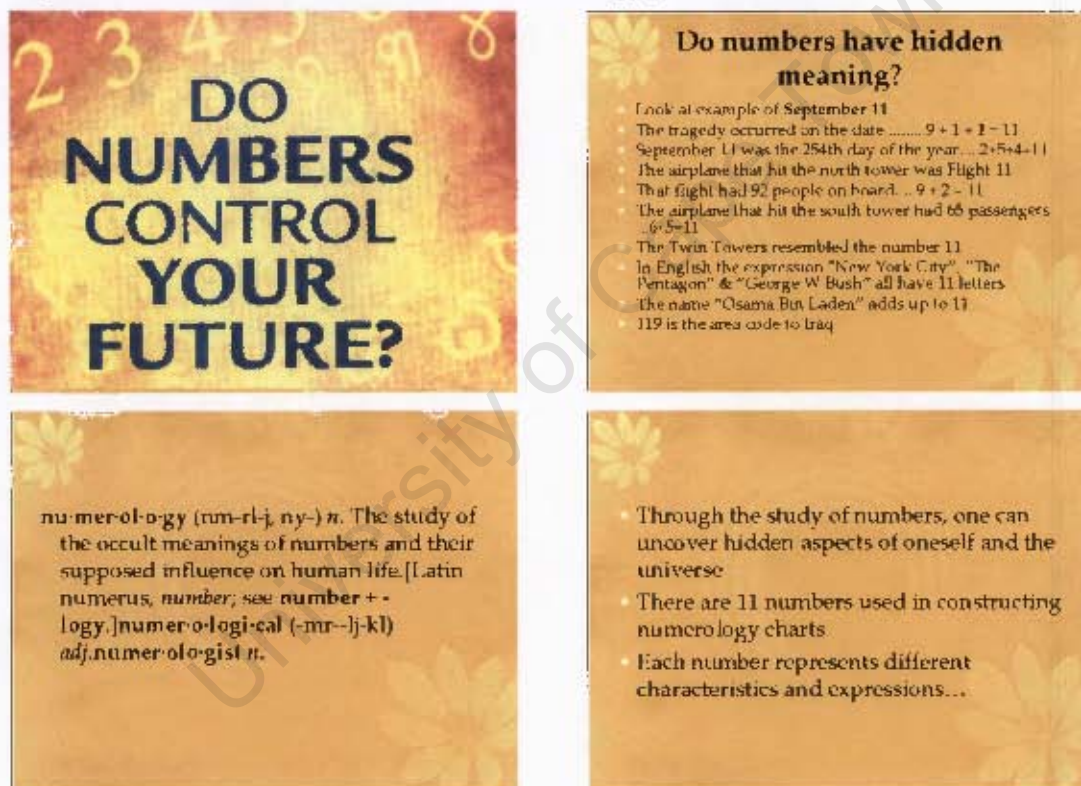
Materials

The staged event took place in a small lecture room in the Psychology Department at UCT. The PowerPoint slideshow designed by the experimenter

⁵ Ethical considerations were reviewed and cleared by the Ethics Board at the University of Cape Town prior to the commencement of this research

for use in the presentation consisted of 44 slides, detailing the various aspects of numerology and each number and letter of the alphabet used in numerology calculations. The script to be used by the perpetrators was also designed by the experimenter and correlated with the slides of the presentation. The entire presentation was designed to last approximately 20 minutes in duration. Both the script and information contained in the presentation was well researched and summarised to be as genuine and believable as possible (Both can be viewed in the Appendix on compact disc).

Figure 11: The first four slides of the numerology presentation



Procedure

The event

The event entailed a live staged incident where participants met with a numerologist (actress) who claimed to be able to read their numerology charts. Participants were exposed to the perpetrator for approximately twenty minutes. The reason for choosing a fairly lengthy event to be witnessed is that this type of situation occurs frequently in reality, such as hostage situations or rape, not only events where the witness sees the perpetrator only

momentarily (for example, a car hijacking or bag snatching). This event was designed to replicate a non-threatening situation where a witness does not have the foreknowledge that memorising the perpetrator's face will be important later, for example, when a con-artist enters one's home and scams one into buying a fake product and then disappears with one's money, never to be heard of again. There was no physical threat and one did not realise that the person's face would need to be remembered for later identification. This effect was aimed for in this study, hence the reason for not debriefing the participants until the reconstruction phase. The three perpetrators had to be convincing and seem authentic to the participants, therefore various websites and books were consulted in order to equip them sufficiently with knowledge on the subject of numerology. Drama students were used as perpetrators as they would be more convincing and able to adapt to the 'role' of a numerologist. They were given a detailed script to rehearse, as well as a PowerPoint presentation (both created by the experimenter) which they would use in their talk.

Participants were asked to meet in a lecture room at the University of Cape Town to attend a presentation on numerology and have their numerology charts prepared. The meeting was about 20 minutes long, during which the numerologist gave them a brief introduction to the study of numerology, with an overview of the basic terms and concepts underlying numerology. At the conclusion of the presentation, she promised to compile a very basic assessment for each individual and asked them to sign up for a follow-up session. They were told that this follow-up session would comprise a one-to-one interview (between each student and the numerologist) regarding their numerology charts. At the following meeting 5 weeks later, the numerologist did not arrive. At this point the experimenter stepped in and informed the participant that the woman was in fact only an actress and that the true reason for their participation was to construct a face composite of the woman using the ID or FACES system, depending on which group they were assigned to. Debriefing involved as much detail of the project as required by the witness.

Stage 3: Reconstruction and morphing phase

Method

Participants

Sixty four of the original 72 participants who witnessed the event took part in the reconstruction phase, as those participants who were familiar with or knew the perpetrator were excluded so as not to bias the results. Twenty four participants had attended the event on day 1; with 20 participants each attending on days 2 and 3. The gender ratio was now equally split, with 32 males and 32 females. Participants used one of two composite systems during all conditions of the reconstruction phase, i.e. ID or FACES. Each of the 'witnesses' in both reconstruction groups were randomly assigned to one of three conditions, namely in-view, from memory with a South African Police Service (SAPS) interview, and from memory with a guided memory interview. The type of condition and use of the 2 composite systems were counterbalanced across participants.

Participants assigned to the in-view condition reconstructed the face with a photograph of the person in view, therefore evaluating the system's basic ability under ideal conditions for amateur users. Participants assigned to the SAPS condition reconstructed the face from memory after being given a South African Police Service (SAPS) interview. Participants assigned to the guided memory condition reconstructed the face from memory after completing a guided memory interview. The latter two conditions evaluate the practical ability of the system relative to a forensic context. In addition to this, an expert user of each of the composite systems (ID and FACES) was asked to do a reconstruction of each of the three perpetrators, therefore evaluating the system's ability under completely optimal conditions and can reveal effects of practice or experience using the system.

The second part of the study tested whether morphing composites produced by multiple witnesses of a target lead to greater identification than multiple

composites used individually. Therefore, the 20 composites reconstructed by the participants assigned to the guided memory conditions for both ID and FACES were used, morphing each set of composites pertaining to each of the three target faces, resulting in 6 additional composites to be evaluated (3 ID 'morphs' and 3 FACES 'morphs').

Materials

One of the composite system's six databases, namely the white female database, was used. The target faces used during the reconstruction phase were of three white female university drama students from UCT. The drama campus is separate to the humanities campus and therefore decreased the chance that their faces would be familiar or known to those students participating in the study. The reason for choosing female perpetrators for this experiment is that it would be more believable and realistic, as it is far more likely that a numerologist would be female than male. It would also be the first time that the female ID database was tested.

Colour photographs were taken of the three perpetrators with a digital camera following the same procedure as mentioned above for the database building, and printed onto high quality photo-paper with a DeskJet printer, each measuring approximately 7.5 x 5 cm. These photographs were used for the in-view condition. The target faces were displayed with a neutral background as a head and shoulder front view of the face with an unsmiling, neutral expression.

The trial face photographs used by participants in the reconstruction phase during their trial run with the respective computer system were of four randomly chosen female faces from the ID white female database. The images of the three perpetrators and four trial faces can be viewed in Appendix B.

Photo lineups for the post-experimental evaluations contained the target face and five distracter faces placed in a table with three rows and two columns,

with a corresponding number, ranging from 1 to 6, at the start of each row. Each cell in the lineup contained 3 different views of each face – a frontal, ¾ and profile view – with each photograph measuring 6 x 8 cm. The reason for including 3 views instead of the traditional frontal view only is motivated by findings in Shapiro and Penrod's (1986) meta-analysis of facial identification studies that showed that faces in three-quarter poses were easier to recognize than faces in frontal view as they display more external features like hair and face shape, which is prominent in unfamiliar face memory (Hancock, 2002). Thus, Shapiro and Penrod's results showed that a three-quarter pose led to the best performance, followed by frontal pose, and then by profile. Using frontal and profile views in the same lineup should provide witnesses with more visual cues to aid them in their identification decisions.

In order to view the entire lineup at once, the rows (27 x 9 cm each) were pasted onto black A3 cardboard sheets. Two lineups were created for each perpetrator, each containing the same six faces, but randomly varying in placement order, totalling six lineups in all (see Figure 12 for an example). All faces were displayed with a black background as a head and shoulder view of the face with an unsmiling, neutral expression.

Figure 12: A 3-view lineup for perpetrator one (lineup 1)



The foils for all the lineups used in this study were chosen to be the most similar-looking faces to the target face out of approximately 220 possible photographs. The foils for the lineups were chosen using the guidelines stated by Malpass, Tredoux and McQuiston-Surrett (2007). Fillers (or foils) resembled the target in significant features, the lineup conformed to the minimum number of fillers (5) per identification procedure, and the target was randomly positioned within the lineup. Lineups were also tested for lineup fairness and effective size using the calculations stipulated in Malpass's Lineup Evaluation "Do-It-Yourself Kit", which can be found on the Eyewitness Identification Research Laboratory, University of Texas at El Paso website (<http://eyewitness.utep.edu/diy.html>). *Lineup bias* refers to whether or not a person unfamiliar with the suspect chooses the suspect from a lineup with a greater or lesser than chance expectancy or guessing, based on the number of members in the lineup (Malpass et al., 2007). This was tested in a mock witness evaluation by asking several people unfamiliar with the suspect to view each lineup and choose randomly who they believed the suspect to be out of the six faces (therefore chance levels are 1 in 6 or 16.67%). *Nominal size* refers to the number of members in the lineup – in this case, the nominal size is six. *Effective size* refers to the number of credible members in the lineup. Lineup bias and effective size for the six lineups can be viewed in Table 2 below:

Table 2: Lineup fairness indices for each of the six lineups used in the evaluation tasks

	Lineup bias	Chance expectancy (1 in 6)	Lineup size (max=6, min=1)
Lineup 1a	0.042	0.167	3.89
Lineup 1b	0.130	0.167	4.37
Lineup 2a	0.000	0.167	3.25
Lineup 2b	0.174	0.167	4.60
Lineup 3a	0.217	0.167	3.70
Lineup 3b	0.087	0.167	4.77

One can conclude from these results that the lineups were fair with an effective size ranging from 3.25-4.77. This means that between 3 and 5

members out of the six-person lineup were credible in terms of similarity to the other lineup members. The 'b' order lineups were not biased, while the 'a' order lineups were biased either towards or away from the perpetrator. This is strange seeing as the exact same lineup members were used for both orders, with the only difference being placement of members within the lineup. However the lineup effective size is still considered adequate, even when considering the lowest value of 3.25.

A standard police interview consists of asking the witness for a cued verbal description of the incident. The investigator prompts the witness by asking about the perpetrator's age, sex, nationality, height, build, facial features, hair, distinguishing marks or characteristics, facial hair, complexion, language, accent, jewellery, accessories (for example, sunglasses, hat or scarf), weapons and clothing. Once all this information is obtained, the expert operator asks the witness to visualise the perpetrator's face, and then proceeds in constructing the facial composite of the perpetrator with the assistance of the witness (Finger & Pezdek, 1999; Prag, 2005). For an example of a detailed script for this type of interview, see Finger and Pezdek (1999).

The SAPS interview used in this study was in the form of an A4 document, which participants were required to complete to the best of their ability, writing the answers in the spaces provided on the form. This form replicates the case report used by the South African Police when opening a criminal case. The document asks participants for a free written description of the suspect(s), after which a cued recall description prompts the witness for specific details, such as age, height, build, individual facial features (nose, lips, chin etc.), facial or body peculiarities and habits. The document can be found in Appendix C.

Development of the guided memory interview for this study

As a result of the poor findings found in my Honours research, I decided to re-evaluate the validity of the guided memory interview designed in 2004. The guided memory procedure is complex, and it is difficult to precisely identify its important characteristics. The guided memory techniques are also a fairly new and under-researched area of study and it is difficult to make assumptions regarding their validity as one does not know if the interview contained too many or too few contextual cues or if length of the interview had any effect. Another factor that may have had an effect is the non-verbal aspect of answering the questions. If participants' minds are allowed to wander, they will generate more incorrect details from imagined memories and commitment effects could occur (Finger & Pezdek, 1999). Participants may have remembered a false memory and committed themselves to its authenticity and in so doing, adhered with this memory even if false. This danger in fabricating memories needs to be well controlled for, by not letting the witnesses' minds wonder during the interview. This was difficult, especially since the 2004 guided memory interview was non-verbal, where the witness visualised the answers to the questions in their minds which were completely out of the control of the experimenter. For this reason the idea of manually writing down the answers to each question was born. Forcing participants to write the answers down on paper could be monitored by the experimenter and would ensure that participants would focus on the task at hand and not allow their thoughts to wonder.

Much discussion was held amongst various members of the ACSENT laboratory about whether non-verbally answering the questions was in fact a good thing to do. How does the questioner know when the participant is finished thinking about the answer? How long should the questioner wait before asking the next question? Therefore, to examine this notion, the 2004 guided memory interview was re-tested using myself as the 'participant' and was adapted to fit an event experienced by myself and another member of the

ACSENT laboratory several weeks previously. The same procedure used in 2004 was applied and the results were valuable.

I discovered that it was actually very difficult to concentrate and think of the answer to one question as the next question was asked soon after, interrupting my thoughts and causing confusion. I found that the guided memory interview was of little beneficial use in its current state and needed serious revisions. Participants were asked not to answer the questions of the 2004 guided memory interview so as to avoid verbal overshadowing (Dodson et al., 1997). However, Meissner and Brigham (2001a) found that the effects are small and can be eliminated by inserting a delay between witnessing and reconstruction. Therefore, the current delay of 5 weeks is sufficient to remove the overshadowing effect and hence witnesses were asked to write the answers to the questions in the spaces provided on the forms. It has also been found by Meissner, Brigham and Kelley (2001) that the verbal overshadowing effect does not occur when a warning is issued to provide only the information of which subjects are absolutely certain, therefore such a warning was asserted at the very beginning of the guided memory interview: "If you are unsure about anything, DO NOT try to answer it." This was again emphasised verbally by the experimenter before participants were allowed to start. I believe that the task of having to fill in the answer to each question forces the witness to really think about what is being asked, instead of thoughts of one question being interrupted by another question posed by the interviewer and ultimately leaving the witness confused and their thoughts jumbled. This method also allows the witness unlimited time to complete the questions. After much thought and discussion with members of the ACSENT laboratory, I decided to make the guided memory interview a written cued recall questionnaire, as opposed to the standard interview procedure where the experimenter asks the participant questions which they must think about and answer in their minds.

This new interview format was then tested in a pilot study. Those participants who had witnessed the event but were unable to continue in the study because they recognised or knew the numerologist (N=8) were asked to

return for their follow-up session 2 weeks after the event. At this session they were debriefed and asked to complete the guided memory pilot interview (in place of creating a composite). Once this was complete, I questioned each participant regarding the usefulness of the questions asked, whether the participant thought anything needed to be added or removed and whether they believed the interview really did assist their memory for the event. Participants were paid R50 for their time and assistance and asked to sign confidentiality forms to not disclose anything relating to the experiment to anyone, as the other participants were still to return for their follow-up session and were unaware of the true nature of the study. Comments from participants included mostly positive points, for example, "nice to use examples", "The slides helped a lot. I had forgotten parts and the slides helped refresh my memory", "questions on the others helped me remember". A few phrases were corrected to be less confusing and more specific, such as question 8 was changed from "Where in the room were the others sitting?" to "Where in the room were the other participants sitting?". The final edited version of the guided memory interview was used for this study.

The guided memory interview is based on the studies by Malpass and Devine (1981), Cutler and Penrod (1988), Cutler et al. (1982), and Krafka and Penrod (1985). Physical, environmental and emotional cues have been adapted from these guided memory interviews to suit the particular situation of this study. The participant was reminded of the events surrounding the meeting of the woman. The participant's memory for the details of the events leading up to the meeting was also explored. Environmental context information was included in the interview as Smith (1979) found that it was a source of useful retrieval cues to aid in recalling information learnt in a specific context. They were asked to silently recall details such as how they felt that day prior to the meeting, where in the room they were sitting, the room's characteristics and so on, and then asked to recall the meeting itself (Gibling & Davies, 1988). No explicit mention was made of the target's appearance throughout questioning, however questions regarding the context of her appearance were asked to evoke memories of the perpetrator's appearance without actually stating any details about her appearance. For example, question 19 states: "Picture her

behaviour. What did she do? Visualise where in the room she was standing and her posture/attitude.”, and question 20 states: “Imagine her facial features. Think about her lips as she spoke, her tone of voice, her eyes during the encounter.”

Krafka and Penrod (1985) used physical contextual cues in their study where accurate identifications almost doubled, therefore the slides shown during the presentation were printed out and given to the witness to aid their memory, as well as the confidentiality form they completed on the day of the event. The contents of the guided memory interview are limited to information about the context of the offence and offender’s behaviour, therefore providing little possibility for biasing witnesses’ subsequent identifications (Malpass & Devine, 1981). Participants were asked to write the answer to the questions in the spaces provided on the form as well as to silently imagine and visualise the event in their minds while doing so.

An advantage the guided memory holds over the standard police interview is that it allows witnesses to describe the events freely in their minds without the constraints of closed-ended questions (Finger & Pezdek, 1999). Malpass and Devine (1981) found that context reinstatement enhanced identification accuracy up to five months after a ‘crime’. The current study tested the delay of 5 weeks, well within the time limit. Examples of both the SAPS interview and guided memory interview questions that were used for this study can be found in the Appendix on compact disc.

In addition to the guided memory interview, a set of guided memory test questions was devised to test participants’ memory of the event, and whether the guided memory interview *improved* the GM participants’ memory for the event and perpetrator. It consisted of multiple-choice questions relating to purely objective and measurable aspects from the guided memory interview that could be directly scored and marked for accuracy. For example, question 4 states:

- 4) Was there any equipment in the room?
- a) Yes, there was an overhead projector and screen
 - b) Yes, there was an overhead projector, screen and laptop (*the correct answer*)
 - c) Yes, there was a laptop and overhead projector
 - d) No, just tables and chairs

Another example, question 7 states:

- 7) Choose the option that **best** describes the information that was portrayed in the slides:
- a) Run-through of the numbers and their meanings as well as each letter of the alphabet
 - b) Run-through of the letters of the alphabet and how the numbers influence the letters
 - c) Explanation of the numbers used in numerology and how to use them to calculate certain 'urges'
 - d) Run-through of the numbers and their meanings as well as the meanings of each letter of the alphabet and how to use these to calculate certain 'urges' (*the correct answer*)
 - e) Detailed example of how to use the numbers and letters and brief explanation of these

These test questions were given to every participant once the reconstruction was complete. An example of the full set of questions can be found in the Appendix on compact disc.

Procedure

Reconstruction phase

At the 'follow-up' meeting, the in-view and SAP participants were individually interviewed in a computer laboratory separate to the room of the event. The same room in which the event was staged was used as the interview room for the Guided Memory participants because studies have shown that

maintaining the same context from study to test increases the probability of recognising a face (Winograd & Rivers-Bulkeley, 1977). There were three instructional manipulations, namely in-view, SAPS interview, and guided memory interview, as well as two system manipulations, namely ID or FACES. Participants were delegated to each condition by random assignment.

The in-view participants were debriefed and told about their real part in the experiment. They were then given a trial run with the aid of the experimenter using either the ID or FACES system, depending as to which condition they were assigned. The experimenter guided the participant through the process of reconstructing a trial face, demonstrating the system's various functions and allowing participants to familiarise themselves with the particular system. Participants were advised not to look at individual features as such, but to view the face as a whole when making decisions. Once confident about using the system, the participants were given a photograph of the perpetrator, and were allowed to proceed with reconstructing the target face with a photograph in view.

The SAPS participants were debriefed and told about their true role in the experiment. Participants were taken to the computer laboratory and asked to complete a standard SAPS interview form, after which they were asked to reconstruct the target face from memory. Again, the participants were given a trial run using the system to which they were assigned with the aid of the experimenter, and once confident, began the target reconstruction from memory.

The guided memory participants again were debriefed. The experimenter then asked participants to complete a guided memory interview questionnaire, and subsequently asked them to reconstruct the target face from memory using one of three laptop computers set up in the 'event' room. Once more, the participants were given a trial run with the help of the experimenter and once confident, they were allowed to begin the reconstruction of the target face from memory. The experimenter instructed the witness to "refresh" their

mental image of the perpetrator's face several times during the composite production process and to refer to the slides as cues in order to minimise or prevent disruption of the mental image (caused by exposure to the facial composite system).

Participants were allowed to stop once they were satisfied they could not produce a better composite, and that their composite represented a good enough likeness to the target face. Although no specific time limit was set, no participant spent more than 80 minutes completing any one composite. Participants took on average 30 minutes to complete the reconstruction. Lack of appropriate female hairstyles in the FACES database proved to be a problem during reconstruction. Therefore, a more fitting hairstyle was chosen from a database of hairstyles (in a similar fashion to the PRO-fit system). The composite, along with the selected hairstyle, were exported to Adobe Photoshop CS Version 8 for modification, and configured to fit the face according to instructions from the participant. The final products of the reconstruction phase were used for the evaluation phase.

Figure 13: Examples of FACES composites before (left) and after (right) editing the hair style in Adobe Photoshop



On completion of the reconstruction the participants were paid R50 for their time and effort and asked to sign confidentiality forms to not disclose anything relating to the experiment to anyone. Participants were also asked to complete a brief post-experimental questionnaire to assess the believability of the event and to test for contamination effects. Participants had to say whether the perpetrator was present in a six-person lineup and if yes, whom they believed the perpetrator was by selecting the number corresponding with the lineup face (The in-view participants viewed the lineup before reconstruction commenced, before they could view the perpetrator's face). They were also asked to comment on the similarity of lineup faces and the believability of the event. Their answers and comments were noted by the experimenter on paper. In addition, each participant was given the guided memory multiple-choice test in order to test their memory for the event. They were not allowed to complete the test in the 'event' room as some questions would be easily answered, and I also did not want the context of the room to have any sort of influence on the results of the in-view and SAPS participants.

In addition to the resulting 64 composites, an expert user of each of the composite systems reconstructed a composite of each of the three perpetrators, producing six additional composites. The resulting composites can be viewed in Appendix B.

Morphing

For the second part of the study, the 20 guided memory composites created by those participants assigned to the guided memory condition for both ID and FACES (11 for FACES and 9 for ID) were used to create 'morphs'. The FaceWarp programme used earlier to create the ID face database was used to create the morphs. Each guided memory composite was individually landmarked in the same manner as discussed previously, and then the group of individual composites pertaining to one perpetrator was inputted into FaceWarp, which created an average of the group of faces inputted (see figures 14 and 15). This average face was, in effect, a morph of the individual composites created for one perpetrator. This averaging procedure filters out

the dissimilar aspects of the individual faces and emphasises the similar aspects, combining the memories of the individual witnesses. This process was repeated for each of the three target faces and for each composite system, only morphing same system composites together. This resulted in 3 'morphs' for each system, therefore producing an additional 6 composites to be evaluated (76 in total). The resulting morphed composites can be viewed in Appendix B.

Figure 14: Three individual ID guided memory composites (top), the morph (bottom left) and the perpetrator (bottom right)

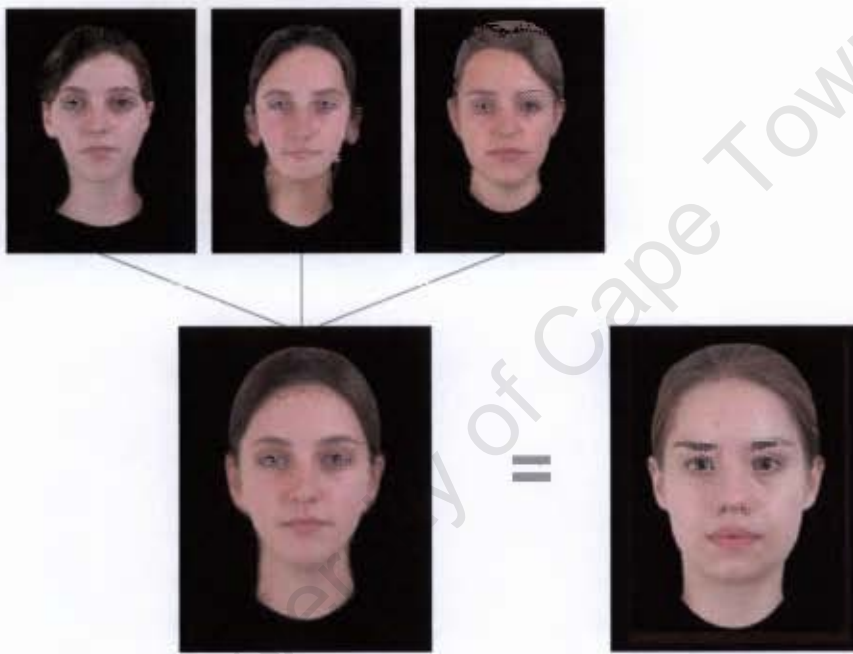
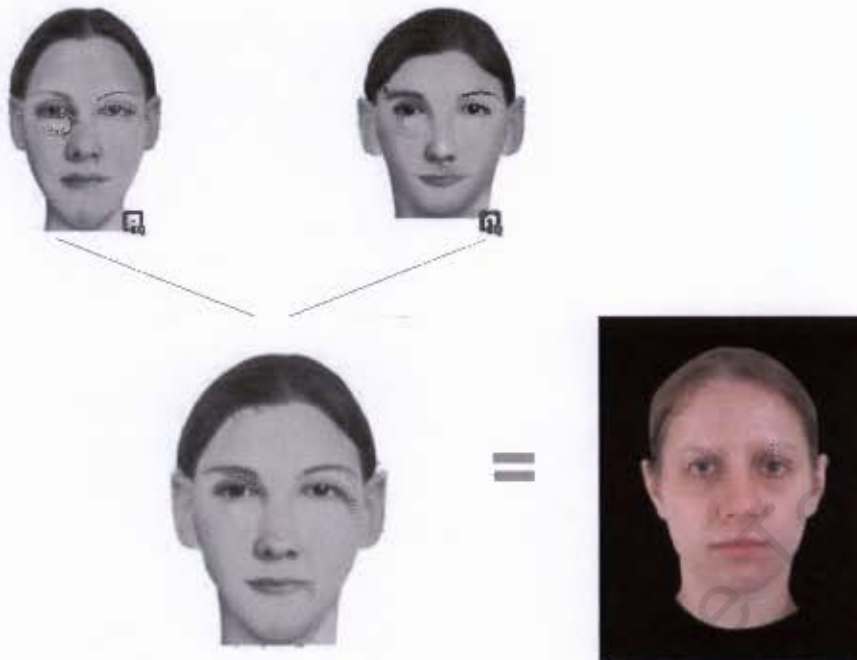


Figure 15: Two individual FACES guided memory composites (top), the morph (bottom left) and the perpetrator (bottom right)



Stage 4: Evaluation phase

Method

Participants

Following the reconstruction phase, 2973 volunteer judges were approached via email to evaluate the reconstructions (a list of convenient emails most likely to be university students and still in use generated in a separate study was used). The members of the photo-lineups were all of UCT students, therefore non-UCT students were approached to complete the evaluations so that there were no biased selections. 2973 non-UCT students were emailed, inviting them to participate in the evaluation and provided a link to the relevant website where the evaluation tasks were located. In thanks for their time and effort, their contact details would be put into a raffle draw to win an Apple iPod. Ultimately, 504 judges evaluated the quality of the 76 composites using rating and matching tasks. Ages of the judges ranged from 18-62 years old,

with an average age of 27.95 years (SD=9.45). Race also varied among judges, with 341 white, 105 black, 36 Indian, 11 coloured, and 8 other (2 did not specify their race).

Materials

The most popular method of evaluating composites produced by participants during the reconstruction phase is the matching task, where participants must identify the closest possible match to the composite. Another popular method, the similarity rating task, requires participants to rate the composites according to the degree of similarity to its target face. Matching tasks are often used together with rating tasks when evaluating composites (For example, Brace et al., 2006; Bruce et al., 2002; Davies et al., 2000; Kovera et al., 1997; Prag, 2005; Sullivan, 2004; Tredoux et al., 1999; Wolgater & Marwitz, 1991), and were both utilised in this study.

The online evaluations consisted of 6 tasks. Each judge was shown a matching and rating task for each perpetrator, with a composite of the relevant perpetrator placed within the 2 tasks. Order of tasks (rating or matching first), composite used and order of perpetrator (e.g. Judge 1 may see perpetrator 2 first, perpetrator 1 second and perpetrator 3 last while judge 10 may see perpetrator 1 first, perpetrator 3 second and perpetrator 2 last) was randomly assigned, therefore counterbalancing across all judges to ensure the form of assessment would not interfere with the results (see Figure 16 for an example).

Figure 16: Examples of the evaluation tasks. Top = similarity rating; bottom = lineup identification after a first choice has been selected



For the photo-lineups used in the evaluation tasks, the target face and five other distracter faces were placed in a table with three columns and two rows, with a corresponding number, ranging from 1 to 6, underneath each picture. Two separate lineups were created for each perpetrator, containing the target face and five different foils randomly varying in placement order within the lineup (this was controlled for by a statistical programme which randomly

shuffled faces within each lineup). These lineups were used for both the matching and similarity tasks, which were placed on the university website for evaluators to voluntarily evaluate. They can be found in Appendix D.

Procedure

Evaluation phase

Five hundred and four independent judges evaluated the 76 resulting composites by performing matching and rating tasks. Each judge was given an evaluation set containing six tasks, randomly varying between the two types of tasks and their two different lineups, and among the three perpetrators. The lineup and type of tasks given for each judge was randomly assigned, as well as which of the 76 composites were used in each task. For example, judge 1 may have received perpetrator 1 in lineup task, lineup 2, perpetrator 2 in lineup task, lineup 1 and perpetrator 3 in similarity task, lineup 2, while judge 2 may have received perpetrator 1 in similarity task, lineup 2 and perpetrator 2 in lineup task, lineup 1. This was done so that the form of assessment did not interfere with the results.

The matching task asked judges to choose from a six-person lineup who the composite assigned to the task represented. Once their choice had been recorded, the image was removed and judges were prompted to give their second, third, fourth and fifth choices, each time the chosen image being removed. The similarity rating implemented a 7-point Likert scale, ranging from 1 (not at all similar) to 7 (extremely similar). Judges were asked to rate the similarity of the given composite to each of the six photographs in the lineup.

Chapter 4

Results

Overview

A holistic, recognition-based computerised composite system, namely ID, was compared to a featural composite building system that is currently in use internationally, namely FACES. The comparison aimed to test whether the ID system produces better quality composites to FACES, as it is based on empirical evidence of people's superior ability to recognise a face rather than the recall of features individually, required by featural systems such as FACES. The study hypothesised that ID would perform better, but results suggest that the two systems performed equivalently.

Numerous studies have shown positive results regarding reinstatement of context, and that guided memory techniques lead to greater identification accuracy and have a significant effect on memory and recognition performance. Therefore, this study aimed to assess whether participants given a guided memory interview prior to reconstruction would produce better quality composites than those participants given the standard treatment observed in reality, namely a police interview. Results suggest that the composites created in the guided memory condition were rated higher than the SAP and morphed conditions, however the difference was non-significant. Therefore, the guided memory interview did not have the desired effect of significantly improving participants' memories of the perpetrator such that their composites were better than the composites created by the other participants. To assess the effectiveness of the guided memory interview on participants' memories of the perpetrator, a post-experimental multiple choice test was administered. This test consisted of questions relating to purely objective and measurable aspects from the guided memory interview that could be directly scored and marked for accuracy. Results showed no significant difference between the three conditions, therefore the guided memory interview did not improve the guided memory participants' memories for the perpetrator's face such that the scores were significantly better than those of the other two conditions.

Studies have reported enhanced identification when multiple composites were combined to create a morph. These studies have shown that morphs were rated as more similar to the target face than the individual composites and as good as the best individual composite. Therefore this study aimed to replicate this finding using the guided memory composites. The guided memory composites for each perpetrator were combined to create three ID and three FACES morphs. Contrary to expectations, the morphed composites performed extremely poorly and were rated the worst and identified the least.

The present chapter reports the results found for the various evaluation tasks according to three central questions:

1. Does the guided memory interview improve composite quality?
2. Do the two composite systems differ in performance?
3. Is there an interaction between the construction conditions and composite systems?

Judges evaluated the quality of the 76 composites using rating and matching tasks.

Data coding

Race and gender were recorded for each judge. Each condition was categorised as IV, GM, SAP, E or M for the in-view reconstructions, guided memory reconstructions, reconstructions after police interview, expert and morphed reconstructions respectively. Which composite system was used for each condition was also recorded. Tasks were designated as either L for a lineup task or S for a similarity task. Judges' decisions as to the lineup tasks were coded as H for a hit or correct recognition of the perpetrator, or as F for a foil choice, where they incorrectly chose a distracter foil face and not the perpetrator. The judges' similarity ratings were scored from 1 (not at all similar) to 7 (extremely similar). The rank of judges' choices was also noted for both the matching and rating tasks. For example, judge one may have given the perpetrator a rating of 4 and was also their highest rating of the six faces. Therefore, the rank is noted as 1 as it was their highest rating. If they had also given ratings of 6 and 7 to two other faces, then the rank would be 3,

as it was their third highest rank. For clarity, one can look at the raw data contained in the Appendix on compact disc.

1. Evaluation identifications

The matching task asked judges to identify the closest possible match to the composite from a six-person lineup. Results are reported in terms of identification rates as a percentage of hits or foil choices.

Descriptive statistics showed that an unusually high number of males ($n=474$) responded to the request for online participation than females ($n=24$). Five judges failed to specify their gender. The number of white judges ($n=341$) was also predominantly higher than the other four race groups combined ($n=160$). The frequency tables proved that the evaluation tasks were completely randomised across tasks, composites and perpetrators and therefore had no effect on the subsequent data.

Pearson Chi-square tests of independence were calculated for judges' decisions across condition, race, task order, gender, composites and system, to determine whether these variables are associated or independent of each other.

Effect of Guided Memory on identification rates

The recognition rates overall differed significantly among the five conditions. The highest rate was found for the expert condition (68.16%), followed by in view condition (53.08%). There is nearly a 30% difference between the in-view and guided memory conditions (24.51%). Thus, the trend is clearly in line with the predicted pattern: the composites created with a photograph in view appeared to be the best, followed by the composites constructed from memory. As Table 3 shows, the expert composites appeared to be better than the in-view composites, in line with expectations as the experts were significantly more practiced in using the relative systems. Looking at the from-

memory conditions, the guided memory composites appeared to be the best followed closely by the SAP and morphed composites, but not by much (3.36%), therefore the effect was not formally tested. This finding is in line with the hypothesis that those participants given a guided memory interview prior to reconstruction would produce better likenesses than those given the standard treatment as would be the case in reality (a police interview). Contrary to expectations, the morphed composites were not better than the guided memory composites. However, most of the composites attracted foil choices, no matter the condition under which they were created or system used, with both the ID and FACES systems performing equally.

Table 3: Overall percentages of hits across all three perpetrators. Number of participants is shown in brackets

Expert	In-view	Guided memory	SAP	Morph
68.16	53.08	24.51	21.15	20.83
(87)	(254)	(78)	(95)	(24)

Assessing GM effects across perpetrators

The identification responses were categorised into hits (correct identifications) and foil choices (lineup foils chosen) and are summarised in Table 4. All participants provided sufficient information to be categorised as either a hit or foil choice.

Table 4: Percentage identification rates (based on judges' first choice selections) for each perpetrator across composite conditions. Number of participants is shown in brackets

Composite condition	Perpetrator 1		Perpetrator 2		Perpetrator 3	
	Hits	Foils	Hits	Foils	Hits	Foils
In-view	50.74 (69)	49.26 (67)	55.29 (94)	44.71 (76)	53.22 (91)	46.78 (80)
Guided memory	10.23 (18)	89.77 (158)	33.80 (24)	66.20 (47)	29.51 (36)	70.49 (86)
SAP	5.61 (6)	94.39 (101)	29.94 (53)	70.06 (124)	27.91 (36)	72.09 (93)
Expert	58.82 (20)	41.18 (14)	73.91 (34)	26.09 (12)	71.74 (33)	28.26 (13)
Morph	6.00 (3)	94.00 (47)	30.77 (12)	69.23 (27)	25.71 (9)	74.29 (26)
Total	23.06 (116)	76.94 (387)	43.14 (217)	56.86 (286)	40.76 (205)	59.24 (298)

As can be seen in Table 4, overall hits for perpetrator one = 23.06%; perpetrator two = 43.14%; perpetrator three = 40.76%. These hit rates show that when using fair lineups, where the perpetrator was difficult to identify from among the six lineup members, the composites are leading to correct identifications almost half of the time (except for perpetrator one who seems to have been difficult to reconstruct, hence the low identification rates).

Chi-square results showed that decision was significant for condition across all three perpetrators (perpetrator 1: $\chi^2(4)=126.1208$, perpetrator 2: $\chi^2(4)=45.51913$, perpetrator 3: $\chi^2(4)=47.77664$; $p<0.0001$ for each comparison). As the observed rates show (see Table 4), the overall level of accuracy is very low apart from the expert and in-view composites, which achieved significantly more correct identifications.

Differences between the composite systems

Table 5: Percentage identification rates for each perpetrator across composite systems based on judges' first choice selections. Number of participants is shown in brackets

System	Perpetrator 1		Perpetrator 2		Perpetrator 3	
	Hits	Foils	Hits	Foils	Hits	Foils
Faces	26.20 (87)	73.80 (245)	43.37 (108)	56.63 (141)	27.89 (53)	72.11 (137)
ID	16.96 (29)	83.04 (142)	42.91 (109)	57.09 (145)	48.56 (152)	51.44 (161)
Total	23.06 (116)	76.94 (387)	43.14 (217)	56.86 (286)	40.76 (205)	59.24 (298)

The chi-square results also showed that decision was significant for system for perpetrators one ($\chi^2(1)=5.437734$; $p=0.02$) and three ($\chi^2(1)=20.91579$; $p<0.0001$), but not for perpetrator two ($\chi^2(1)=0.0108516$; $p=0.92$) (see Table 5). Both FACES and ID appeared to perform poorly, but ID seemed to perform marginally better than FACES. However, the difference was small (approximately 4% - see Table 6) and non-significant:

Table 6: Overall percentages of hits across all three perpetrators. Number of participants is shown in brackets

FACES	ID
32.49 (248)	36.14 (290)

Interaction between the conditions and systems

Figure 17: Interaction identification rates for perpetrator one

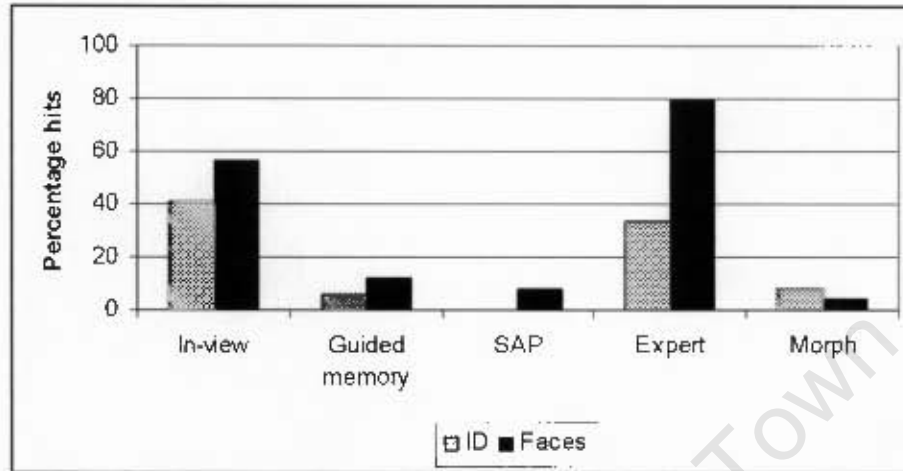


Figure 18: Interaction identification rates for perpetrator two

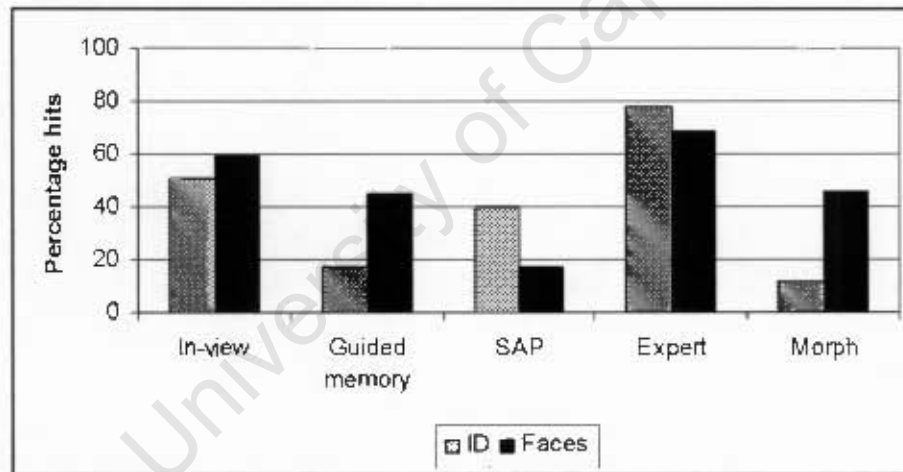
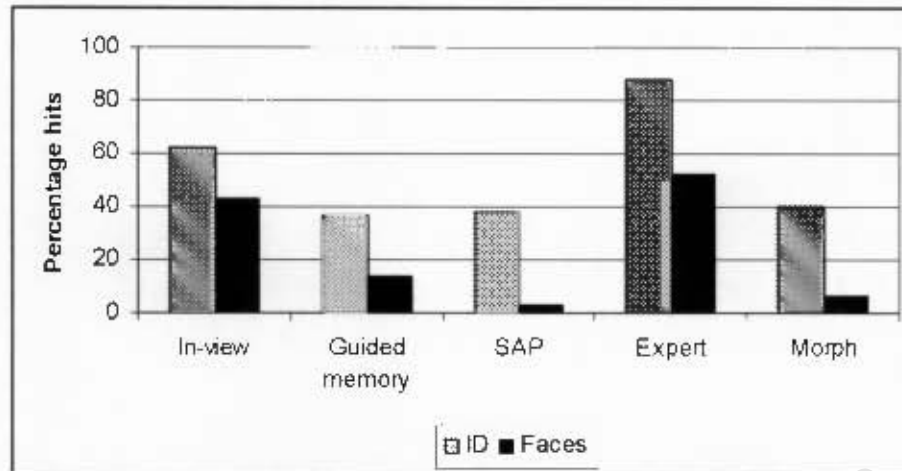


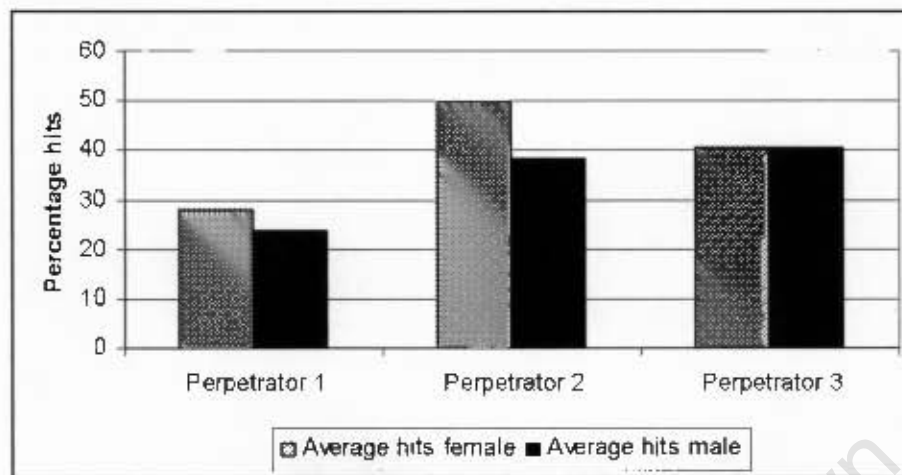
Figure 19: Interaction identification rates for perpetrator three



The figures above (Figures 17, 18 and 19) illustrate the interaction effects between construction condition and composite system for each of the three perpetrators. All three interactions were significant (perpetrator one: $\chi^2(9)=141.9132$, perpetrator two: $\chi^2(9)=66.69305$, perpetrator three: $\chi^2(9)=82.78660$; $p<0.0001$ for each comparison). It seems certain participants were better at reconstructing the perpetrators than others, and this also varied according to the composite system used. Inferring from Figures 17, 18 and 19, certain perpetrators seemed easier to reconstruct using one system than the other under the various conditions. ID appeared to perform better for perpetrator three for all conditions. Generally FACES was better for perpetrator one for all conditions except the morphs.

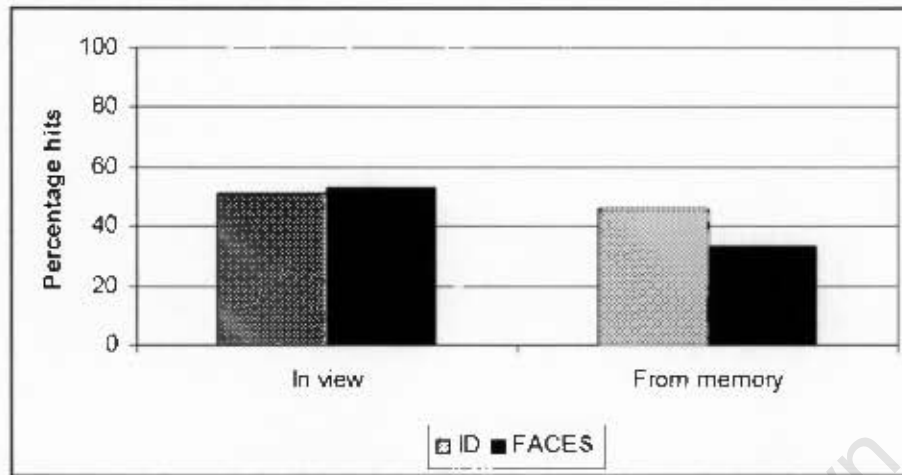
McKelvie (1981) found a gender interaction for face recognition, with an own-sex advantage for females. Gender effects could also have occurred here, where the female participants may be better able to reconstruct the perpetrator faces than males, particularly since the target is their own sex. As can be seen in Figure 20, the female participants constructed better likenesses than their male counterparts, corresponding with the findings of McKelvie (1981).

Figure 20: Percentage identification rates for each perpetrator across gender of reconstructor



Frowd, Carson, Ness, McQuiston-Surret, et al. (2005) believe that after a delay, the memory of a face is more of a holistic impression or gestalt than a set of individual features, and it is this shift in processing strategy that they believe facilitates the use of a holistic over a featural-based construction system. The in-view and from-memory conditions were therefore compared to assess this view that a configural system is better for constructing composites from memory and featural systems are better for composites constructed with the target face present. The SAP and guided memory conditions were combined to form the 'from-memory' condition. The expert and morph conditions were dropped from this analysis as they were not part of the original reconstruction conditions, therefore only the in view composites constituted the 'in-view' condition. Looking at Figure 21, one can clearly see that ID was better in the 'from-memory' condition, while FACES was marginally better for the 'in-view' condition. However these effects do not seem to be significant and were not tested. A possible reason for the small difference in identification rates between ID and FACES in the 'in-view' condition may be the newly included 'optimise features' tool in the ID system, which allows one to manipulate individual features within the context of a whole face.

Figure 21: Identification rates for the ID and FACES systems across perpetrators for two memory conditions



Gender, race and task order effects on identification rates

Chi-square results showed that decision was not significant for race across all three perpetrators and therefore was dropped from further analyses. Gender and task order were also found to be non-significant and did not have an effect on the results, and were also dropped from further analyses (refer to Table 7).

Table 7: Pearson Chi-square test of contingency results.

	Perpetrator	Chi-square	df	p
Condition x decision	P1	126.1208	4	0.0001
	P2	45.51913	4	0.0001
	P3	47.77664	4	0.0001
Race x decision	P1	0.659896	4	0.95618
	P2	4.797313	4	0.30874
	P3	1.749948	4	0.78163
Task order x decision	P1	0.146901	1	0.70152
	P2	0.325532	1	0.56830
	P3	0.659503	1	0.41674
Gender x decision	P1	0.523861	1	0.46920
	P2	0.308079	1	0.57886
	P3	0.002624	1	0.95915
Composite x decision	P1	203.5948	27	0.0001
	P2	103.5199	23	0.0001
	P3	116.2792	23	0.0001
System x decision	P1	5.437734	1	0.01971
	P2	0.010852	1	0.91703
	P3	20.91579	1	0.00001

The results show that the composites for each of the perpetrators were significantly associated with the decision made, namely a hit or foil choice ($p < 0.0001$ for each comparison). The identification rates for each of the 76 composites constructed were obtained in order to assess their effectiveness and accuracy (recall that 28 composites were created for perpetrator one, 24 for perpetrator two and 24 for perpetrator three). For perpetrator one, nearly all composites (3, 5, 9-20, 22, 23, 24 and 27) attracted foil choices by judges, with only composites 1, 7 and 26 attracting a higher than chance amount of correct choices ($\chi^2(27) = 203.5948$). Composites 3, 10, 11, 14, 16, 19, 20, 23 and 24 in fact received no hits at all. Results for perpetrator two showed an

even spread of hits (12 out of 24) and foil choices (14 out of 24) amongst the composites, with only composite 18 obtaining no hits ($\chi^2(23)=103.5199$). For perpetrator three, 10 out of 24 composites attracted correct identifications while 14 out of 24 composites attracted foil choices ($\chi^2(23)=116.2792$). Only composite 20 received no hits.

Rates of choosing of lineup members

Chi-square goodness of fit analyses were also calculated in order to determine whether judges selected the perpetrator faces from the lineups at greater than chance levels or purely by random guessing. These goodness of fit tests compared matching performance (in terms of identification rates) to chance expectation (1 in 6). If the perpetrator was chosen randomly by someone unfamiliar with the perpetrator out of a lineup of six faces, the chance that the perpetrator would be chosen is one out of six ($1/6$). In this case, once the first choice was made, that lineup member was removed, leaving five faces from which to select their second choice. Now the chances that the perpetrator is chosen (if not already chosen) is two out of six ($2/6$). This process was repeated after each selection until all members of the lineup had been selected, with each selection increasing the chances that the perpetrator would be chosen. These chance levels were used as expected frequencies and were compared to the observed frequencies. The observed frequencies were calculated by cumulatively adding the number of first, second, third, fourth and fifth choices to compare against the expected frequencies of $2/6$, $3/6$, $4/6$ and $5/6$ respectively. Hasel and Wells (2005) found a prototype effect, where morphs came to resemble the non-target faces more than the individual composites, and it seems this is may have happened in this study. For clarity please refer to Table 8 below. Looking at the data per choice (1-6), the observed frequencies are hoped to be higher than the expected (chance) frequencies and therefore demonstrating deliberate choosing as opposed to randomly picking anyone within the lineup.

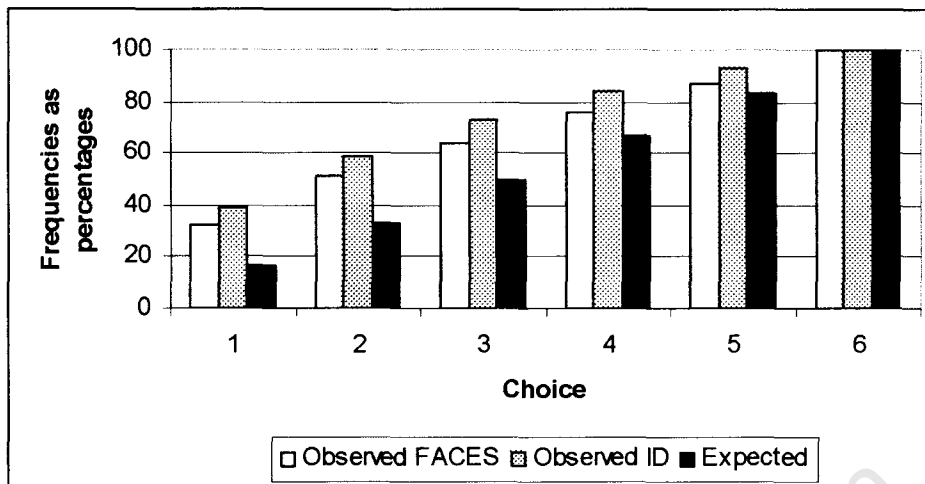
Table 8: Expected and observed frequencies for each perpetrator. Note: Observed = Raw cumulated frequencies

Choice	Perpetrator 1			Perpetrator 2			Perpetrator 3		
	Raw	Observed	Expected	Raw	Observed	Expected	Raw	Observed	Expected
1	32	32	28.5	110	110	42.5	153	153	52.2
2	31	63	57	50	160	85	66	219	104.3
3	21	84	85.5	42	202	127.5	42	261	156.5
4	19	103	114	19	221	170	29	290	208.7
5	33	136	142.5	17	238	212.5	18	308	260.8
6	35	171	171	17	255	255	5	313	313

Separate chi-squares were conducted for each perpetrator, evaluating judges' decision for each choice (first, second, third, fourth or fifth selection from the lineup) for system and condition. Results showed that identification rates were significantly better than chance for all three perpetrators (perpetrator 1: $\chi^2(5)=15.66899$; $p=0.008$, perpetrator 2: $\chi^2(5)=464.4547$; $p<0.000001$, perpetrator 3: $\chi^2(5)=432.1078$; $p<0.000001$). Thus the perpetrator was not chosen randomly from the lineup, showing that other factors had influenced the judges' choices when selecting a face. The lineups have already been proven to be fair and unbiased towards the perpetrator, therefore one could assume that the composite aided in their choice.

Overall, both systems were found to be significant ($p<0.0001$ for both), therefore the type of composite shown in the evaluation task aided in judges' selection of the perpetrator from a six-person lineup. Looking at the rates of guessing collapsed across perpetrators (see Figure 22), both ID and FACES seemed to be chosen at above chance levels for each choice (except the last choice which obviously would be equal to chance).

Figure 22: Rates of choosing collapsed across perpetrators for FACES and ID



Condition was also found to be significant, except for the morphed composites (see Table 9). This implies that the morphed composites were not good likenesses as the perpetrators were chosen randomly when morphed composites were shown.

Table 9: Rates of choosing collapsed across perpetrators for the five construction conditions. Obs = observed (raw cumulated) frequencies; Exp = expected frequencies

Choice	Expert		Guided Memory		In-view		Morphed		SAP	
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
1	86	21	78	61.5	254	79.5	24	20.7	96	68.8
2	104	42	147	123	353	159	41	41.3	180	137.7
3	116	63	205	184.5	408	238.5	59	62	244	206.5
4	125	84	256	246	441	318	81	82.7	300	275.3
5	125	105	316	307.5	462	397.5	100	103.3	359	344.2
6	126	126	369	369	477	477	124	124	413	413

Conclusion

Overall, the perpetrators were chosen purposefully from the lineup, and that factors such as the condition under which the composite was constructed and the composite system used had influenced the judges' choices when selecting a face. The 'from-memory' composites were constructed far better when using

the ID system, while the 'in-view' composites were constructed better when using FACES but this advantage over ID was small. Over all choices made by the judges (1-6), both FACES and ID performed poorly, with ID showing a marginal (but non-significant) advantage over FACES.

The lineup task aimed to test whether the composite was a good-enough likeness to aid an unfamiliar person in identifying the perpetrator, for example, when a composite is broadcast on television and a person recognises the likeness as a neighbour. The rating task aimed to assess whether the composite is a good quality representation or likeness of the perpetrator, in terms of similarity to a perpetrator's photograph. Tasks are related such that poor ratings should equate with more foil identifications. The evaluation ratings are presented in the next section.

2. Evaluation ratings

The similarity rating task required participants to rate the composites according to the degree of similarity to each member of a six-person lineup, using a 7-point scale. Results are reported in terms of ratings as a number ranging from 1 to 7.

Race and gender effects on similarity ratings

Due to the nature of the evaluation tasks, possible cross-race and cross-gender effects may occur. Namely, white judges would be expected to be superior to the 'not white' judges in their comparative ability for the perpetrator and lineup foils as all six members were white. Female judges would also be expected to give more accurate ratings than male judges as the lineup members were female. In order to assess whether judges' race and/or gender had an effect on the similarity ratings, two-way repeated measures ANOVAs (between-factors, repeated on factor B) were conducted, comparing similarity ratings across race and gender. Table 10 provides all cell means and

standard deviations. The number of Indian, 'coloured' and 'other' race groups were very low ($n = 36, 11$ and 7 respectively), therefore race was collapsed into two groups, namely 'white' and 'not white'.

Table 10: Comparative similarity ratings of composites across perpetrators and race of judges (a = Mean; b = Std. Dev.; c = n)

		Perpetrator 1	Perpetrator 2	Perpetrator 3	Marg. Means (race)
White	a	3.548	3.956	4.003	3.836
	b	1.866	1.847	1.941	
	c	341	341	341	
Not white	a	3.453	3.610	4.346	3.803
	b	1.964	1.993	2.038	
	c	159	159	159	
Marg. Means (perpetrator)		3.501	3.783	4.174	

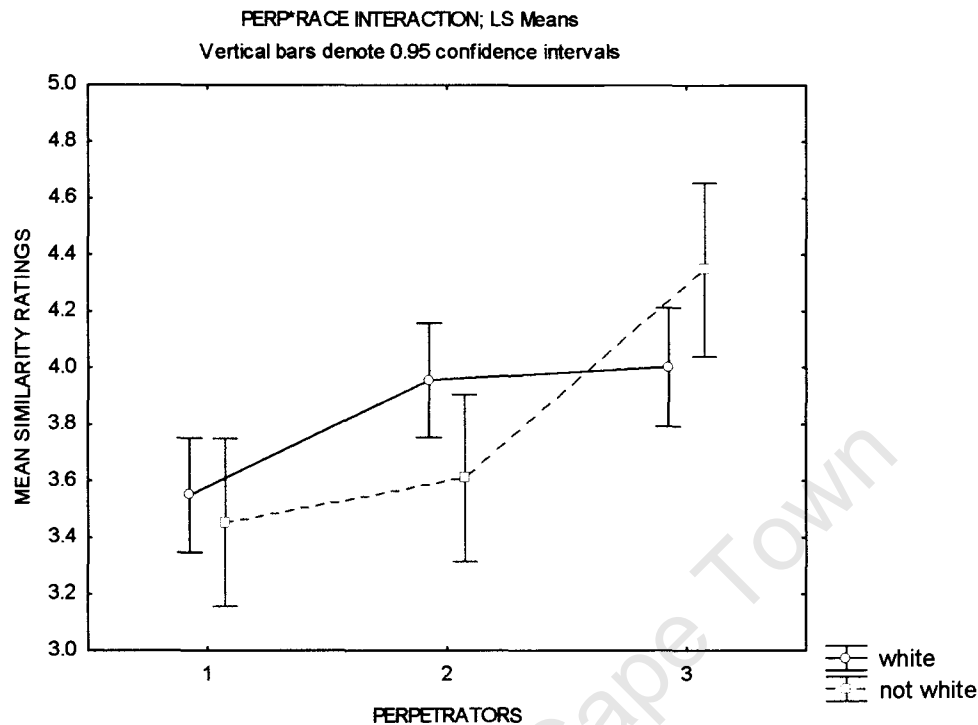
A significant interaction effect was found between race and perpetrator ratings, $F(2, 990)=4.344$; $p<0.05$, as can be seen in Table 11.

Table 11: ANOVA summary table

Source	SS	df	MS	F	p
<u>Between Subjects</u>					
Race	0.35	1	0.35	0.070	0.791402
Error (w/in grps)	2495.22	498	5.01		
<u>Within Subjects</u>					
Perpetrator	99.33	2	49.66	16.358	0.000001
Perpetrator*race interaction	26.37	2	13.19	4.344	0.013237
Error (B x Subjects w/in grps)	3023.76	996	3.04		

Tukey's HSD comparisons were conducted to assess where the significance lay. Looking at Figure 23, there seems to be clear perpetrator effects, with an overall tendency for perpetrator three to be rated highest ($M=4.11$) followed by perpetrator two ($M=3.85$) and perpetrator one ($M=3.52$) across both race groups. As can be seen by the non-overlapping confidence intervals, white judges rated perpetrators two and three significantly higher than perpetrator one, while the 'not white' judges rated perpetrator three significantly higher than perpetrators one and two.

Figure 23: Similarity ratings across the three perpetrators for both race groups. Note: W = white; NW = not white



No clear cut cross-race effects was found, hence the variability in the results can be attributed to effects regarding the different perpetrators. Perpetrator effects are common in this type of study, where the three females used as perpetrators were chosen randomly, and distinctiveness and memorability of their faces could not be pre-determined. Some faces are just more memorable or easier to reconstruct than others (for example, Green & Geiselman, 1989; Laughery and Fowler, 1980) and it seems this was the case here. Certain reconstructions were clearly better than others leading to differing ratings across perpetrators as well as race groups.

Similar results were found for gender, with a significant interaction effect found between gender and perpetrator ratings, $F(2, 990)=3.0798$; $p<0.05$ (see Table 13). Table 12 provides all cell means and standard deviations.

Table 12: Comparative similarity ratings of composites across perpetrators and gender of judges (a = Mean; b = Std. Dev.; c = n)

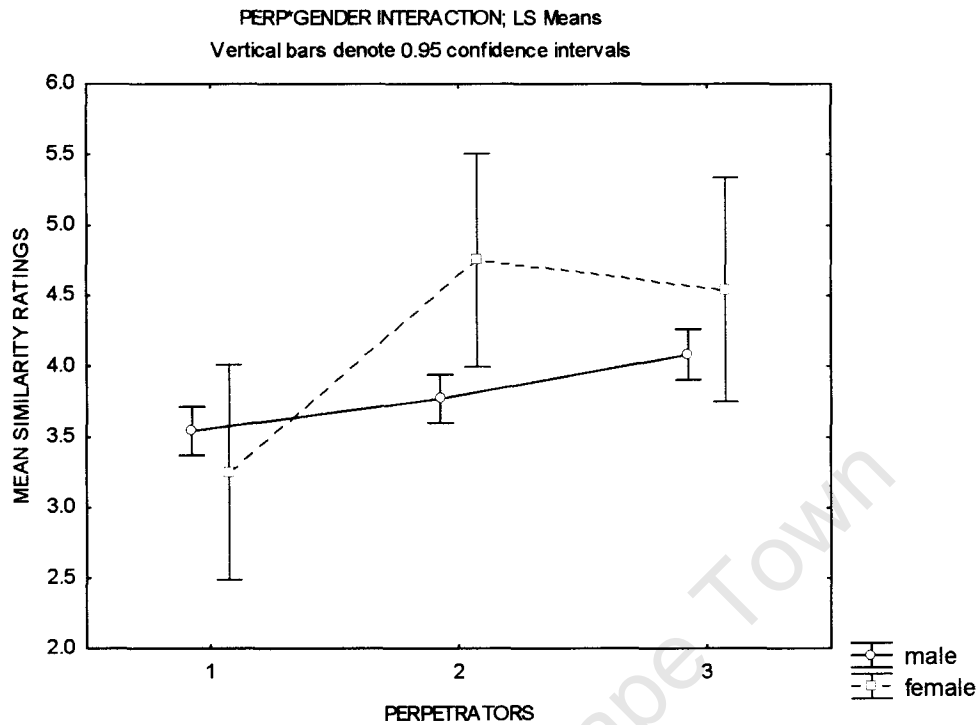
		Perpetrator 1	Perpetrator 2	Perpetrator 3	Marg. Means (gender)
Male	a	3.541	3.770	4.082	3.798
	b	1.916	1.904	1.981	
	c	473.000	473.000	473.000	
Female	a	3.250	4.750	4.542	4.181
	b	1.482	1.391	1.956	
	c	24.000	24.000	24.000	
Marg. Means (perpetrator)		3.396	4.260	4.312	

Table 13: ANOVA summary table

Source	SS	df	MS	F	p
<u>Between Subjects</u>					
Gender	10.042	1	10.042	2.0046	0.157451
Error (w/in grps)	2479.606	495	5.009		
<u>Within Subjects</u>					
Perpetrator	48.404	2	24.202	7.9855	0.000363
Perpetrator*gender interaction	18.669	2	9.334	3.0798	0.046407
Error (B x Subjects w/in grps)	3000.465	990	3.031		

Tukey's HSD comparisons were conducted, and again it seems the significance can be attributed to perpetrator effects as there was no observable cross-gender effect (see Figure 24). It seems that certain perpetrators were rated higher or lower than others, depending on the judge's gender. There was an overall tendency for perpetrators two ($M=3.82$) and three ($M=4.10$) to be rated significantly higher than perpetrator one ($M=3.53$), as can be seen by looking at the non-overlapping confidence intervals. Male judges rated perpetrator three significantly higher than perpetrator one, while female judges rated perpetrator two significantly higher than perpetrator one.

Figure 24: Similarity ratings across the three perpetrators for males and females



Interaction between the conditions and systems

Separate two-way repeated measures ANOVAs (between-factors, repeated on both factors) were conducted for each perpetrator, comparing similarity ratings across condition and system. Due to the perpetrator effects found in the previous ANOVA analyses, perpetrators were separated and analysed individually in order to assess whether the same effect would be found again across the conditions and systems.

Perpetrator one

A significant main effect was found for condition and system, however the interaction between condition and system was not found to be significant ($p > 0.05$; see Table 15). Table 14 provides all cell means and standard deviations.

Table 14: ANOVA summary table

Source	SS	df	MS	F	p
Condition	274.254	4	68.563	23.553	0.000001
System	27.217	1	27.217	9.350	0.002351
Condition*system interaction	5.572	4	1.393	0.479	0.751492
Error	1435.110	493	2.911		

Table 15: Comparative similarity ratings of composites across condition and system for perpetrator one (a = Mean; b = Std. Dev.; c = n)

		In-view	Guided memory	SAP	Expert	Morph	Marg. Means (system)
FACES	a	4.652	3.000	3.221	5.609	3.091	3.914
	b	1.712	1.810	1.700	1.559	1.758	
	c	92	139	68	23	11	
ID	a	4.000	2.717	2.629	4.500	2.529	3.275
	b	1.629	1.622	1.610	1.382	1.841	
	c	53	53	35	12	17	
Marg. Means (condition)		4.326	2.858	2.925	5.054	2.810	

However, considering the mean similarity ratings shown in Figures 25 and 26, FACES composites were rated significantly higher ($M=3.91$) than the ID composites ($M=3.28$).

Figure 25: Perpetrator one similarity ratings across conditions for the two systems

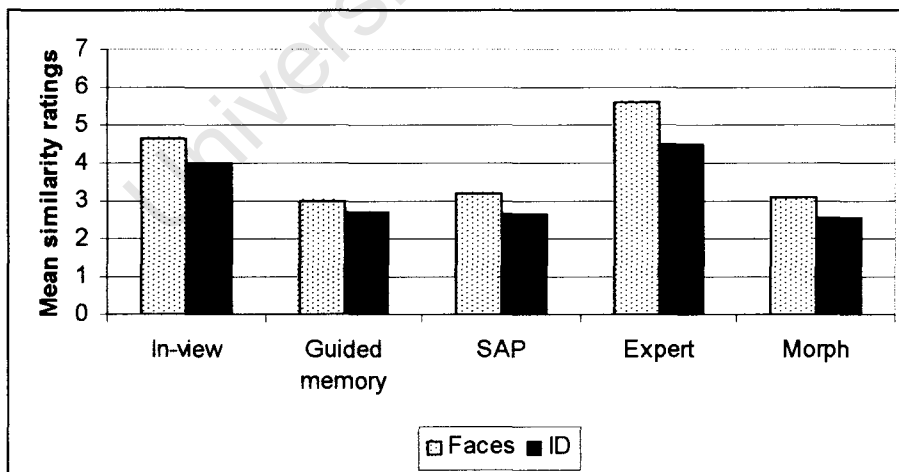
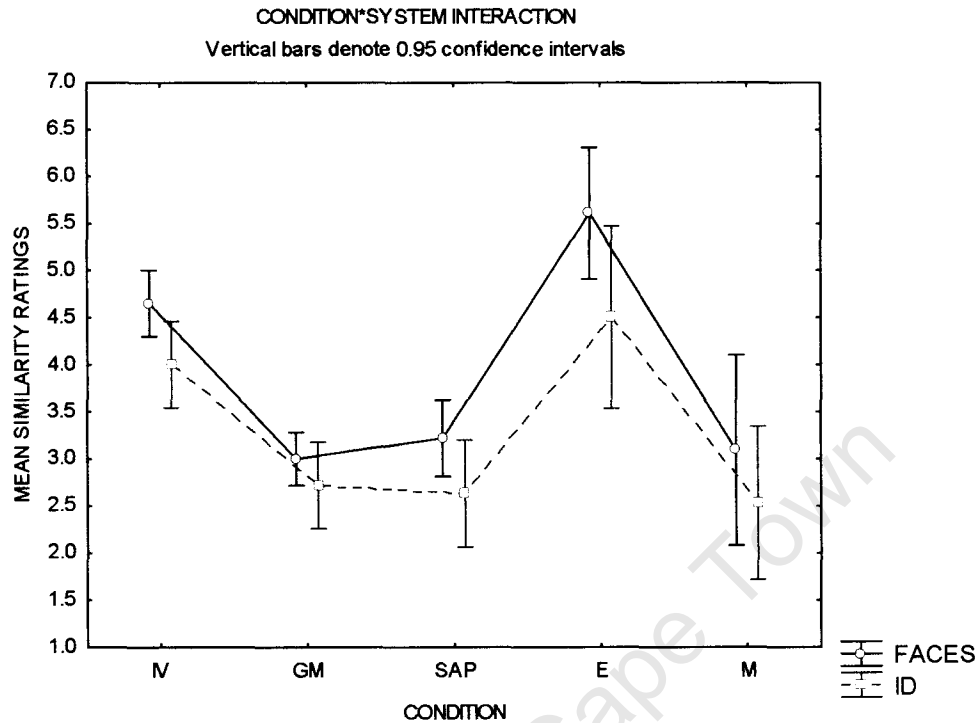


Figure 26: Graph of cell mean profile of similarity ratings for each condition and system for perpetrator one



One can also see that expert composites achieved significantly higher ratings ($M=5.23$) than all the other conditions. This was expected, as the expert reconstructors appear to be far more experienced in creating reconstructions than the participants in this study.

Perpetrator two

Interaction between condition and system was found to be significant, $F(4, 492)=4.883$; $p<0.01$ (see Table 17). In addition, a significant main effect was found for condition. Table 16 provides all means and standard deviations.

Table 16: ANOVA summary table

Source	SS	df	MS	F	p
Condition	177.966	4	44.491	14.223	0.000001
System	3.814	1	3.814	1.219	0.270079
Condition*system interaction	61.102	4	15.275	4.883	0.000720
Error	1539.081	492	3.128		

Table 17: Comparative similarity ratings of composites across condition and system for perpetrator two (a = Mean; b = Std. Dev.; c = n)

		In-view	Guided memory	SAP	Expert	Morph	Marg. Means (system)
FACES	a	5.026	3.792	3.320	4.353	3.692	4.037
	b	1.821	1.812	1.702	2.060	1.715	
	c	77	53	75	17	26	
ID	a	3.810	3.303	3.240	5.727	3.050	3.826
	b	1.888	1.667	1.718	1.667	1.504	
	c	79	33	100	22	20	
Marg. Means (condition)		4.418	3.548	3.280	5.040	3.371	

Tukey's HSD comparisons were conducted across condition and system to determine where the significance lay (see Figures 27 and 28). With regards to the FACES system, the in-view composites achieved significantly higher ratings (M=5.03) than the guided memory (M=3.79), SAP (M=3.32) and morphed composites (M=3.69). With regards to the ID system, the expert composites achieved significantly higher ratings (M=5.73) than all the other conditions.

Figure 27: Graph of cell mean profile of similarity ratings for each condition and system for perpetrator two

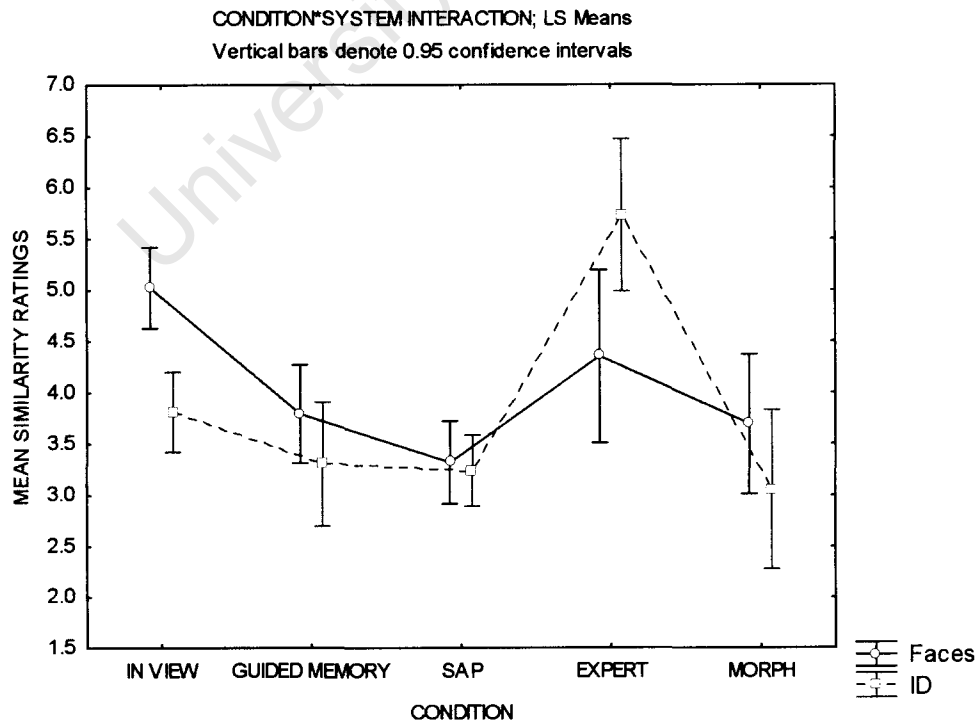
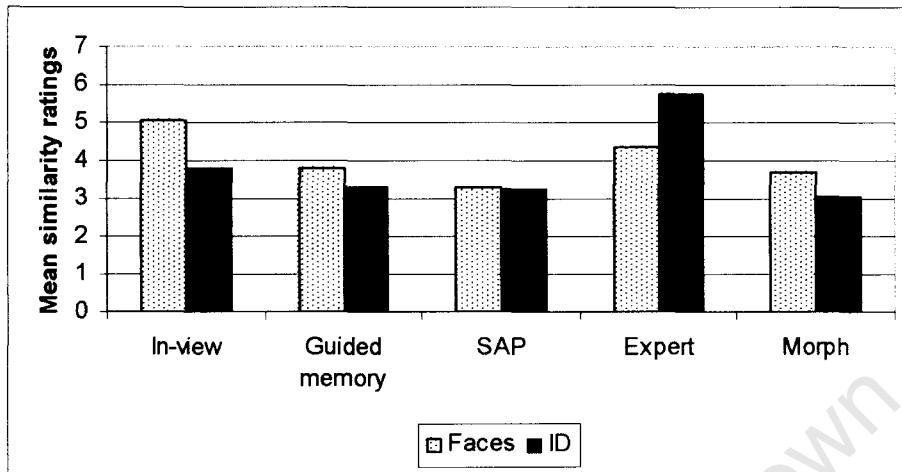


Figure 28: Perpetrator two similarity ratings across conditions for the two systems



Perpetrator three

A significant main effect was found for condition only. The main effect for system as well as the interaction between condition and system was not found to be significant ($p > 0.05$; see Table 18). Table 19 provides all cell means and standard deviations.

Table 18: ANOVA summary table

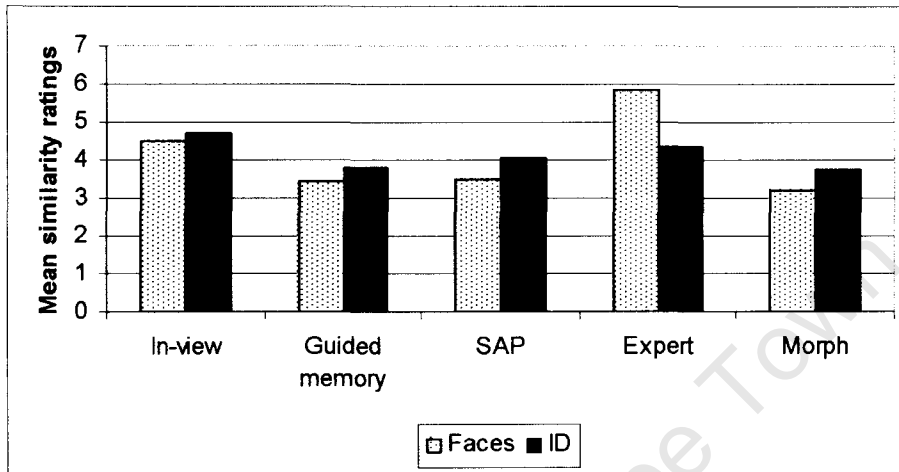
Source	SS	df	MS	F	p
Condition	133.864	4	33.466	9.219	0.000001
System	0.028	1	0.028	0.008	0.930571
Condition*system interaction	32.810	4	8.203	2.260	0.061740
Error	1789.675	493	3.630		

Table 19: Comparative similarity ratings of composites across condition and system for perpetrator three (a = Mean; b = Std. Dev.; c = n)

		In-view	Guided memory	SAP	Expert	Morph	Marg. Means (system)
FACES	a	4.483	3.442	3.520	5.864	3.222	4.106
	b	1.823	1.906	1.717	1.859	1.826	
	c	87	43	50	22	27	
ID	a	4.679	3.807	4.031	4.353	3.750	4.124
	b	1.915	1.959	2.000	2.149	2.049	
	c	84	88	65	17	20	
Marg. Means (condition)		4.581	3.624	3.775	5.108	3.486	

However, considering the mean similarity ratings shown in Figure 29, FACES and ID composites seemed to achieve like similarity ratings across conditions (FACES M=4.10; ID M=4.12).

Figure 29: Perpetrator three similarity ratings across conditions for the two systems



Conclusion

Overall, results suggest that the composites created in the guided memory condition were rated higher than the SAP and morphed conditions, however the difference was non-significant. Therefore, the guided memory interview did not have the desired effect of significantly improving participants' memories of the perpetrator such that their composites were better than the composites created by the other participants. To assess the effectiveness of the guided memory interview on participants' memories of the perpetrator, a post-experimental multiple choice test was administered. This test consisted of questions relating to purely objective and measurable aspects from the guided memory interview that could be directly scored and marked for accuracy.

3. Evaluation of the guided memory interview

Section 3 reports the findings for the guided memory post-experimental multiple choice test, comparing the test scores across the three construction conditions, as well as comparing the two composite systems' performance. Results are reported as average scores for each condition or system as a result of one-way analyses of variance.

Did the guided memory interview improve participants' memory significantly more than those participants not given an interview?

Simple one-way ANOVAs were conducted comparing overall performance (test scores) across the three composite construction conditions, namely in-view, guided memory and SAPS. Table 20 provides all cell means and standard deviations.

Table 20: Means and standard deviations for construction conditions for Study 1. (gm=guided memory; iv=in-view; sap=police interview)

	Condition		
	gm	iv	sap
Mean	9.077	8.889	8.842
St. Dev.	1.801	1.745	1.385
n	13	18	19

Results showed that condition did not have a significant effect on test scores ($p = 0.92$; see Table 21). Looking at Table 20 one can see that the guided memory participants did score higher than the in-view and SAPS participants, but the difference was not significant.

Table 21: ANOVA summary table

Source	SS	df	MS	F	p
Condition	0.453	2	0.226	0.085	0.918674
Error	125.227	47	2.664		
Total	125.680	49			

4. Familiarity and lineup identifications resulting from the post-experimental questioning

Lastly, section 4 examines the results found for the post-experimental questionnaires administered to the composite reconstructors after completion of the composites, and is reported qualitatively in terms of participants' subjective feedback and opinions. The post-experimental questionnaire tested the memory of each participant for the event and face, and the post-experimental lineup choice assessed whether contamination had occurred from using the two composite construction systems.

Identification of the perpetrator from a six-person, 3-view lineup

The post-experimental questionnaire results show that in total, 96% of participants correctly identified the perpetrators, with only 2 false choices made (and one, who had initially said the perpetrator was not present in the lineup, chose the perpetrator when forced to choose). Therefore almost every participant had encoded the face in memory sufficiently to recognise it quite a time (5 weeks) later, demonstrating that memory was not observably contaminated by the reconstruction task.

Believability of the study in terms of participants' personal feedback

Most participants noted that the event was extremely believable and "fooled them completely". They had no idea that the study truly involved face recognition and some were even disappointed that they would not receive their numerology chart! Several participants said that they were suspicious of what was going on, but had no idea that the study was related to face recognition whatsoever. One participant commented: "It was very suspicious for the psychology department to be testing numerology, but I didn't know it was on face recognition." Ultimately all participants had no idea that the true reason for their participation was to remember and reconstruct a face.

When asked about the similarity of the faces within the various line-ups, participants gave mixed reactions. Most thought that the members of the lineups were similar to each other, with others saying the complete opposite, that they were "Quite different to each other." Many participants commented that she stood out because they "remembered her face so well" and that they "recognised her immediately". It can be said, therefore, that the event duration of 20 minutes was more than sufficient for the participants to encode the perpetrator's face into memory. One participant even commented that she "remember(ed) her nose and ears so the $\frac{3}{4}$ view really helped. I probably wouldn't have chosen her if I didn't have that view". Another participant commented that "She looks similar to 5 but not to the others, her profile is very different", leading me to believe that the 3 different facial views did aid the identification process. It can also be said from these results that using the composite systems and viewing multiple different facial images did not contaminate the participants' memory for the witnessed face.

Many participants, however, commented on the difficulty in reconstructing the perpetrator's face using the two systems. Many admitted that their composite did not resemble their memory of the perpetrator, but they just could not get a better likeness.

Chapter 5

Discussion

Overall, the results of this study were relatively poor. Identification rates were poor when composites were reconstructed from memory and judges tended to make more foil choices than identifying the perpetrator. Likewise, the similarity ratings were concentrated around the lower end of the 7-point scale, with the highest average rating being 5.23 (for the expert composites), but most falling around 3.

Three hypotheses were tested in this study:

1. Does ID produce better quality likenesses than FACES?
2. Does context reinstatement improve memory of a face such that the resulting composite is an improvement on those constructed from memory alone?
3. Do morphs of multiple composites from different participants produce a 'super composite' that is better than the average individual composite?

Each of these will now be discussed, comparing the results of this study to others that have also looked at these issues.

1. Does ID produce better quality composites than FACES?

ID appeared to perform marginally better than FACES, but the difference was small (approximately 4%) and insignificant. It seems that FACES has an advantage when composites are constructed in-view, however consider that featural manipulators are particularly well suited to reconstructions made in full view, as specific features can be compared and matched to those on the visible target. Prag (2005) found that the FACES system was better for reconstructing composites in-view, while ID was better for reconstructing composites from memory, which leads to the notion of a feature matching

strategy when using a photograph to reconstruct a face. A possible reason for the similarity in performance of ID and FACES in this study may be due to the recent addition of the 'optimise feature' tool that allows users to alter individual features within the context of a face, which is suggestive of a dual processing strategy as opposed to a purely holistic strategy. FACES presented both a complete face as well as a set of isolated features from which to select, while ID only presented a complete face from which features could be exchanged.

A featurally distinctive face may be better reconstructed when the construction system is capable of focusing on specific features (like FACES and other featural composite systems), whereas a more typical face may be better reconstructed using a more holistically-oriented composite building strategy such as that employed by ID (Koehn et al., 1999). Evidence for this hypothesis can be seen in the results for Study 1. FACES composites seemed to be better for perpetrator one while ID composites seemed to be better for perpetrator three. If one refers to Figure 30, one can tentatively assume from the photographs that perpetrator one is fairly distinctive, with memorable features such as her eyebrows and lips, while perpetrators two and three are fairly average-looking. Therefore, the featural system appeared to perform better when a distinct face was being constructed and the holistic system appeared to perform better for an average face. Green and Geiselman (1989) also found that distinct composites were better recognised than composites of average-looking faces.

It is likely that faces vary in terms of ease or difficulty to reconstruct and likewise, that people differ in their ability to reconstruct faces (Laughery and Fowler, 1980). This could explain the variability of the results obtained for the study. Another factor that could have influenced the results is task demand. The participants did not know what procedure to expect and had to become familiar with the system and task demands in a relatively short time.

Results of a study by Wolgater and Marwitz (1991) showed that familiarisation and practice with the composite system led to better composites (both in-view and from-memory) of higher accuracy. As can be seen by the results, practice

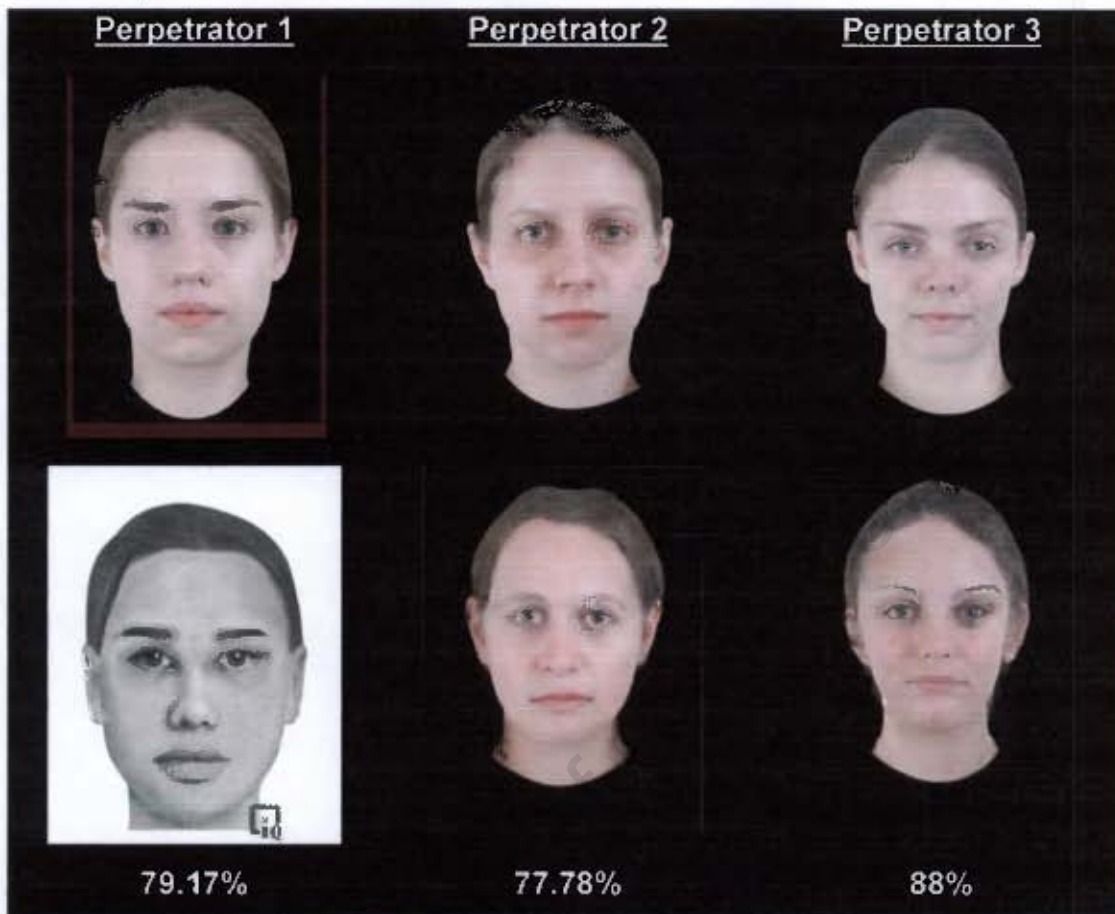
using the systems does lead to better composites, as the expert condition was rated best for all three perpetrators. Therefore, it may be useful to make real eyewitnesses practice using the system before reconstructing the witnessed criminal.

A possible source of inconsistency in quality among the composites is the ease with which different faces are remembered. For example, Shepherd and Ellis (1973, cited in Ellis, 1975) showed that both attractive and unattractive girls' faces are better recognised than those of average attractiveness. Perpetrator effects are common in this type of study, where the three females used as perpetrators were chosen randomly, and distinctiveness and memorability of their faces could not be pre-determined. Some faces are more memorable or easier to reconstruct than others (Ellis, 1975; Green & Geiselman, 1989; Laughery and Fowler, 1980) and it seems this was the case here. Certain reconstructions were clearly better than others leading to differing ratings across perpetrators as well as race groups and gender. In order to minimise this problem, the ID white female database needs to be expanded so that it is more representative of all types of faces, particularly ones with distinctive features. The same can be said for FACES, whose feature libraries could be updated and expanded to include more typical and distinctive features.

Another notable point is that people clearly differ in their ability to reconstruct faces (Ellis et al., 1975; Laughery & Fowler, 1980). The variation in accuracy of the composites produced indicates that certain participants may have been better at reconstructing faces than others (as was found in a study by Green and Geiselman, 1989), and may also have been dependent on which composite system was used. Boice, Hanley, Shaughnessy and Gansler (1982) believed that eyewitness identification is an observational skill similar to that demanded in other situations. They found that subjects who were superior in other general observational tests tended to be superior at recalling the details of a witnessed crime scene. Some achieved quite good likenesses to the original faces, receiving identification rates of up to 88%, while others made very poor reconstructions that received no hits whatsoever.

As can be seen in Figure 30, 2 of the 3 best identified composites were created using ID by the expert user. The best FACES composite was reconstructed for perpetrator one by an in view participant and was closely followed by the expert composite constructed using FACES (78.95%). The second best composite for perpetrator two was constructed using FACES by a guided memory participant and obtained an identification rate of 75% (compared to 77.78% for the best composite which was constructed by the expert user with the ID system). Likewise with perpetrator three, the second best composite followed closely the best composite identification rates, receiving 87.5%, and was constructed by an in view participant using ID. This shows that not only expert operators can achieve high quality likenesses as some participants managed to reconstruct very good composites and were identified often.

Figure 30: The best identified composite (bottom) of each perpetrator (top) with percentage of correct identifications underneath



2. Does context reinstatement improve memory?

The recognition rates overall were fair and differed significantly among the five conditions. The expectation that the composites created with a photograph in view would be better than the composites constructed from memory was confirmed. There is nearly a 30% difference between the in-view and guided memory conditions. As Table 3 (repeated below for convenience) shows, the expert composites appeared to be better than the in-view composites, in line with expectations as the experts were significantly more practiced in using the relative systems. Looking at the from-memory conditions, the guided memory composites appeared to be the best followed closely by the SAP and morphed composites. This is in line with the hypothesis that those participants given a guided memory interview prior to reconstruction would produce better likenesses than those given the standard treatment as would be the case in

reality (a police interview). Contrary to expectations, the morphed composites were not better than the guided memory composites. However, most of the composites attracted foil choices, no matter the condition under which they were created or system used, with both the ID and FACES systems performing equally.

Table 3: Overall percentages of hits across all three perpetrators. Number of participants is shown in brackets

Expert	In-view	Guided memory	SAP	Morph
68.16	53.08	24.51	21.15	20.83
(87)	(254)	(78)	(95)	(24)

If judges made their choices randomly, they would correctly guess the perpetrator once in every six attempts, an accuracy rate of 16.7%. Identification rates above this level indicate that the composites are providing information that is aiding judges in their selections. Therefore identification performance was compared to chance expectation. Results from the chi-square goodness of fits showed that the perpetrators were not chosen randomly from the lineup (those tests were conducted on a sequence of choices, *not* on the data in Table 3), therefore one can assume that the composites aided in their choice. The only condition that was not found to be significant was the morph condition and implies that the morphed composites were not good likenesses as the perpetrators were chosen randomly when they were shown.

The same pattern found for identification rates was observed in the similarity rating data. FACES composites for perpetrator one achieved higher ratings than the ID composites. For perpetrator two, FACES composites were found to be significantly better than ID composites, except for the expert condition, where ID received higher ratings. However, perpetrator three showed that ID and FACES achieved similar ratings across conditions. It seems that FACES performed marginally better, but the evidence is not clear enough to make an overall conclusion regarding which of the two systems is better. The rating pattern across all three perpetrators was also very similar: expert composites

always received the highest ratings followed by the in view composites. This predicted result is consistent with the general findings of several previously conducted experiments evaluating a range of manual and computerised facial composite systems that reconstructions done with the target face present are more accurate than those reconstructed from memory (For example, Davies et al., 2000; Ellis et al., 1975; Frowd, 2001; Tredoux et al., 1999; Tredoux et al., 2006; Wogalter & Marwitz, 1991). The from-memory conditions tended to differ slightly, with SAP composites rated better than guided memory and morphed composites for perpetrator one and three, but guided memory composites rated better than morphed and SAP composites for perpetrator two. These small differences were probably due to perpetrator effects and were not found to be significant.

The results from the post-experimental multiple choice test showed that the guided memory interview implemented in this study did not significantly improve participants' memory. The guided memory participants did appear to score marginally better than the in-view and SAP participants, however the effect was non-significant. Average scores across all conditions were 8.92 (55.77%). Guided memory participants' scores ($M = 9.08$) were above the average for the group, but this advantage boils down to answering one or two more questions correctly than the other conditions. The non-significant results found for the guided memory technique implemented in this study could have resulted from the lengthy delay inserted in this study. However, Finger and Pezdek (1999) found that inserting a delay of 1 hour from viewing a face to recognition significantly increased identification accuracy in the cognitive interview condition from 47% to 85%. Davies, Ellis and Shepherd (1978) found a slight decline in recognition accuracy after a 3-week delay, but generally found facial memory to be quite resistant to decline over time. Retention interval is a strong predictor of recognition accuracy in facial recognition and eyewitness identification experiments. In eyewitness literature, retention intervals vary from less than 1 hour to 5 months. The average retention interval found by Shapiro and Penrod's (1986) meta-analysis was 108 hours (4.5 days). Some may question whether the delay

implemented in this study (5 weeks) was perhaps too long, but then it must be noted that Malpass and Devine (1981) found a remarkable improvement in memory using guided memory techniques 5 *months* after the initial event.

However, at this point more investigation needs to be done. The guided memory questions used in the test need to be examined in more depth and perhaps include a comparison group in subsequent testing who experience an even longer delay, to assess whether the substantial memory improvement found by Malpass and Devine (1981) and Krafka and Penrod (1985) can be replicated using the interview schedule developed for Study 1.

The results for the expert condition of this experiment compared favourably against those of Cutler et al. (1988), where a 49% identification rate was found after getting an experienced operator to make in-view reconstructions of 10 target faces. They also compare well against the results found by Gibling and Bennet (1994), who used experienced Photofit operators and found in-view identification rates of 54%. However, the outcome of Koehn and Fisher's (1996) experiment revealed a better identification rate of 77% when an experienced Mac-a-Mug Pro operator made in-view reconstructions. Similarly, in Tredoux et al.'s (2002) evaluation of ID, an impressive average identification rate of 86% was found when an experienced ID operator was used.

Looking at the overall 'in view' category (combining expert and in view conditions), the identification rates found in this study (60.62%) are quite favourable in comparison to many other studies that have implemented this condition. For example, Davies and Oldman (1999) obtained a meagre 12.5% identification rate for E-fit composites constructed with a photograph of the target present. Brace et al. (2000) and Frowd (2001) also found low identification rates (24.95% and 25.3% respectively) when composites were created in view. Prag (2005) and Tredoux et al. (1999) both obtained identification rates of 51% when composites were reconstructed in view using

the ID system. Study 1's results fair relatively better, with an identification rate of 60.62%. This finding is also a large improvement on the results from my Honours thesis (Sullivan, 2004), where only 33% of in-view composite identifications were correct.

Looking at the overall 'from-memory' category (combining the guided memory, morphed and SAP conditions), results showed a 22.16% identification rate. This result compares favourably to studies conducted by Koehn and Fisher (1996) who found overall Mac-A-Mug Pro likeness ratings to be very poor with only 4% correct identifications from the lineup; Davies and Oldman (1999) whose subjects correctly named only 7.6% of the targets overall when composites were created using E-fit; and Ellis, Shepherd and Davies (1975; 1978a) who found only 12.5% of identifications made by judges of Photofit composites were correct. All the aforementioned studies obtained identification rates of less than 20% after a relatively short delay (Koehn & Fisher, 1986) or no delay at all (Davies & Oldman, 1999; Ellis, Shepherd & Davies, 1975; Ellis, Shepherd & Davies, 1978a), which implies that the results of Study 1 were rather favourable.

However, other studies have found better results, such as Christie et al. (1981) who found 30% correct identifications after a one-minute delay for from-memory composites created using the CADDC system. Tredoux et al. (1999) also found significantly higher rates, obtaining 43% correct identifications. Wogalter and Marwitz (1991) found a matching score of 40% when Mac-a-Mug Pro composites were produced from memory following a brief 8-second exposure. Davies et al. (2000) found a matching accuracy of 63% after a one-minute exposure to the target face, which is much higher than the rates obtained in this study. However, these delays are short in comparison to the 5 week delay experienced in Study 1.

The results of Study 1 are comparable to those found by Christie and Ellis (1981), who obtained an identification rate of 23% when Photofit composites

were constructed from memory immediately following a 1 minute exposure to a target face. Again Study 1 obtained equivalent identification rates after a much longer (5 week) delay, which is impressive. It must be noted, however, that comparisons across experiments are made difficult due to the differences in delay, type of incident (live incident or viewing a photograph or video), and context reinstatement procedures. All of these factors could possibly interfere with the context reinstatement procedures (Cutler, Penrod & Martens, 1987).

Overall guided memory identification rates were 24.51%. These rates are equivalent to those found in my Honours thesis, and compare poorly with Gibling and Davies's (1988) study concerning context reinstatement, where a 49% enhancement of memory was obtained. However when comparing the identification rates to studies that have employed the Cognitive Interview, this study fares better. For example, Frowd et al. (2004) found 8.5% identification rates for EvoFIT composites created from memory and a mere 3.7% for PROfit composites. Frowd, Carson, Ness, Richardson, et al. (2005) again found dismal rates when comparing EVOfit, PROfit and E-fit composites created from memory after being given a Cognitive Interview. They found a 19% identification rate for E-fit, 17% for PROfit and a pathetic 1.5% for EVOfit composites.

Extensive empirical examination, including this study, indicate that composite systems produce poor quality composites when constructed from memory, which are difficult to match to target faces (Christie & Ellis, 1981; Davies et al., 2000; Ellis et al., 1975; Koehn et al., 1999; Koehn & Fisher, 1996; Kovera et al., 1997; Laughery & Fowler, 1980). However, a recent study by Frowd, Carson, Ness, McQuiston-Surrett, et al. (2005) found a 42.1% overall correct identification rate for composites produced from memory after a 2 day delay. They also found that FACES composites produced about 30% correct identifications. These rates are not particularly high, but they are substantially better than the rates found in a previous study by the authors (Frowd et al., 2004).

3. Do morphs produce a 'super composite'?

The results showed effects contrary to those predicted: a morph of individual guided memory composites was not rated better than the individual guided memory composites. In fact, the morphs were rated the poorest of all conditions. As Bruce et al. (2002) have already shown, morphs should be as good as the best individual composite and better than the average individual composite, and these laboratory findings were replicated by Frowd, Bruce, Storås, Spick and Hancock (2006) in a genuine criminal investigation (Operation Mallard). The guided memory composites were not very highly rated and received fairly low identification rates (24.51%); hence the individual composites were probably not very good likenesses to begin with. Therefore morphing these poor likenesses resulted in an even poorer morph which combined incorrect information leading to the opposite of what was expected. Hasel and Wells (2005) found a prototype effect, where morphs came to resemble the non-target faces more than the individual composites, and it seems this is may have happened in this study.

Morphing composites was hoped to improve composite quality as it combines the memories of multiple witnesses. However, if these witnesses all have a poor memory for the target face, then morphing these composites will not lead to a better quality image. In fact, the opposite is bound to happen, which is what was found in this Study. These results challenge Bruce et al's (2002) findings and the external validity of this method. There is no way of assessing whether one composite is better than another prior to a suspect's arrest, therefore morphing composites from multiple witnesses may inadvertently de-emphasise a well-matched feature, leading to the prototype effect mentioned by Hasel and Wells (2005).

Future avenues to explore

Koehn et al. (1999) suggest that in laboratory studies, judges who are familiar with the target face should evaluate the resulting composites and not unfamiliar judges, as composites used in real criminal investigations are most useful when the police officer referring to it is familiar with the target. The

composite may remind a police officer of a particular suspect, or exclude a known suspect from the investigation if there are significant differences.

Both holistic and featural information are likely to contribute to normal face recognition, depending on the nature of the task. People may use a particular strategy that is the most appropriate for a specific task. However, no knowledge exists as to which processing strategy is naturally used by people (Hole, 1994). There is also evidence that different processing strategies may be used for familiar and unfamiliar faces (Hines & Braun, 1990). In order to effectively test this hypothesis, one would need to compare the results found in this study with a replicate study using famous female faces as perpetrators instead of the unknown women used here.

The guided memory procedure uses semantic information about the context of an event as a source of associative cues to facilitate memory for the physical appearance of the suspect. The contents of the guided memory script are limited to information about the context of the event and suspect's behaviour, therefore providing little possibility for biasing witnesses' subsequent identifications. It would be beneficial if police could use context reinstatement to improve identification performances, especially after long delays between a witnessed offence and request for recognition or composite construction (Malpass & Devine, 1981).

Chapter 6

General conclusion

As results from the post-experimental lineup task showed, the perpetrators' faces were remembered well (96% identification accuracy from a six-person lineup), thus one possible explanation for why the composites were so poor could be that the faces were very distinctive, therefore attempting to reconstruct them proved difficult due to a lack of distinctive features available in the FACES system's database. Similarly, the ID system may have lacked sufficient variety of eigenfaces from which to build the perpetrator's face. When participants reconstruct composites from memory, they are struggling to match each feature with one that has to be retrieved from a memory trace undergoing decay from the delay between event and reconstruction (5 weeks) as well as interference from the multiple eigenfaces that are viewed simultaneously (Ellis, Shepherd & Davies, 1975). As Brace et al. (2006) noted, when a person is presented with multiple composites, they may be able to make use of the similarities and differences to extract additional information. However, if too many are shown, the variety is too large or there are too many images to compare simultaneously, and the task becomes too complex. The same problem may have occurred when using ID to reconstruct a composite, as the GUI displays 12 faces at a time. As noted by da Costa (1998) and Davies and Valentine (2007), after many generations during composite production, the faces tend to become very similar to one another, therefore producing composites that are a combination of the target face and other similar foils.

The danger of viewing similar likenesses may lead to interference with memory of the original face to be remembered (Comish, 1987). However, this study showed that 96% of participants correctly identified the perpetrator after completion of a composite, contrary to the danger stipulated by Comish (1987). This excellent identification rate shows that the event was sufficiently

long enough to ensure proper encoding of the face, and has positive implications for real-life situations where the criminal is often seen for longer than 20 minutes.

Mauldin and Laughery (1981) in fact found that composite construction *improved* recognition accuracy such that participants who constructed a composite were extremely more likely to recognise the target in subsequent recognition tasks (90% identification accuracy as opposed to 60% for controls who did not reconstruct a face composite). This seems to have been the case in this study as almost all participants remembered the perpetrator's face after the reconstruction task and after a long delay.

Perhaps the reason for the guided memory interview having such a small effect on memory is that questions asked in the interview were not helpful or did not elicit sufficient contextual cues to aid participants' memory. The average total scores from the post-experimental multiple choice test for the guided memory participants were 8.92 out of 16 (55.77%). Unfortunately no similar tests were conducted pre-reconstruction, therefore it cannot be said whether memory increased after the guided memory interview was administered. This needs to be addressed in future research in context reinstatement. The guided memory composites were not good likenesses, therefore combining them led to even worse quality morphs. If one combines the similarities of poor likenesses, one will most likely end up with an average poor likeness.

The ID composite system under evaluation is constantly being updated and improved. As can be seen in the numerous studies of da Costa (1998), Prag (2000), Tredoux et al. (1999), Tredoux (2001), and Tredoux et al. (2006), much has been improved in the nine years since its inception in 1998, for example, the 'optimise features' tool and new higher quality database images that have been meticulously landmarked with many more points than previous (which gives a much clearer, more human-looking eigenface from which to work with). As emphasised by da Costa (1998), the basic assumptions to this system are revolutionary. Firstly, faces are produced according to a configural

rather than fragmentary process. Secondly, it makes use of people's greater ability to recognise faces as opposed to recalling faces. Thirdly, it eliminates the need for verbal descriptions or expert operators.

Mentally reinstating the context surrounding an event has been shown in numerous laboratory studies to aid memory. This effect has important ramifications as it is easier than returning physically to the crime scene, and may be preferable as the crime scene may change over time and the possible emotional trauma experienced at the event is minimised (Geiselman et al., 1986). Reinstatement of context is a promising factor in the aid and enhancement of eyewitness performance (Cutler, Penrod & Martens, 1982). However, realism of the staged event could affect results. As Yuille (1993) notes, most participants in laboratory studies act as "unaffected bystanders" (p. 572) whereas a real crime setting can vary greatly from a simulated event. Motivation across participants was likely to differ – the participants were aware that they were part of a laboratory experiment and knew that their contribution would have little impact outside the laboratory. The incentive to produce recognisable composites is also much greater in real life: the arrest of a criminal compared with a small monetary reward (Frowd et al., 2005).

A solution to this problem would be to use real witnesses of crime as subjects, but this is very problematic. The researcher cannot tell for certain if the witness is correct in their choices or who the real criminal in fact is (unless the case has been solved and a photograph of the suspect is provided, which can be compared to the reconstruction). For this reason, laboratory studies are conducted to test methods to help in these and other complex tasks.

The guided memory techniques are a fairly new and under-researched area of research and it is difficult to make assumptions regarding results as one does not know if the interview contains too many or too few contextual cues or if length of the interview has any effect. The validity and reliability of the guided memory interview has not been well reported, and which questions to ask to be effective is unknown, therefore it will be difficult to measure the 'real' effectiveness of the interview. Generalisability of the results is also risky since

the study uses fixed effects as opposed to random effects. Fixed effects make use of instances or representations of the population, making generalisability difficult as it is almost impossible to apply the results to the entire population.

Further research in this area will increase our knowledge about how and which contextual cues affect memory, and may help in developing procedures that will increase the reliability of eyewitness identifications (Cutler & Penrod, 1988). More testing needs to be done around this area of research with larger samples and more perpetrators in order to test the validity and reliability of the guided memory techniques. Only three different faces were tested in this study, which may not be representative of other faces (Comish, 1987). Davies and Valentine (2007, p. 33) suggest that

“while measurable progress has been made and all systems may claim successes, perhaps the quest for the perfect system may be illusory and we must learn to live within the limitations of the witness memory.”

Despite the disappointing empirical findings, composites are nevertheless invaluable to police inquests. Even with their shortcomings, they are useful for eliminating unlikely suspects in mug files or lineups prior to exposure to witnesses, narrowing the potential group of suspects and improving chances of identifying and apprehending offenders (Brignull, 1998; Davies & Valentine, 2007; Green & Geiselman, 1989; Prag, 2005). Very often in criminal investigations a composite of a suspect's face is the only possible lead. As Brignull (1998) asserts, until a practical alternative has been developed, these systems are simply the best available.

References

- Barber, P. (1988). *Applied Cognitive Psychology. An information-processing framework*. London: Methuen and Co. Ltd.
- Bekerian, D.A. & Bowers, J.M. (1983). Eyewitness testimony: were we misled? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9, 139-145.
- Bekerian, D.A. & Dennett, J.L. (1993). The cognitive interview technique: reviving the issues. *Applied Cognitive Psychology*, 7, 275-297.
- Bennett, P.J. (1986). Face recall: a police perspective. *Human Learning*, 5, 197-202.
- Boice, R., Hanley, C.P., Shaughnessy, P. & Gansler, D. (1982). Eyewitness accuracy: a general observational skill? *Bulletin of the Psychonomic Society*, 20, 193-195.
- Brace, N., Pike, G., Kemp, R., Turner, J. & Bennett, P. (2006). Does the presentation of multiple facial composites improve suspect identification? *Applied Cognitive Psychology*, 20, 213-226.
- Brace, N.A., Pike, G.E. & Kemp, R.I. (2000). Investigating E-FIT using famous faces. In A. Czerederecka, T. Jaskiewicz-Obydzinska & J. Wojcikiewicz (Eds.). *Forensic Psychology and Law* (pp.272-276). Krakow: Institute of Forensic Research Publishers.
- Bradshaw, J.L. & Wallace, G. (1971). Models for the processing and identification of faces. *Perception and Psychophysics*, 9, 443-448.
- Brigham, J.C., Maas, A., Snyder, L.D. & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology*, 42, 673-681.
- Brignull, H. (1998). Why do face reconstruction systems give such poor results and how might they be improved? Retrieved September 15, 2004, from <http://neocortex.co.uk/Identikit.htm>
- Brown, C. & Lloyd-Jones, T.J. (2002). Verbal overshadowing in a multiple face presentation paradigm: effects of description instruction. *Applied Cognitive Psychology*, 16, 873-885.

- Bruce, V. (1988). *Recognising faces*. Hove, U.K: Lawrence Erlbaum Associates.
- Bruce, V., Ness, H., Hancock, P.J.B., Newman, C. & Rarity, J. (2002). Four heads are better than one: combining face composites yields improvements in face likeness. *Journal of Applied Psychology*, 87, 894-902.
- Cabeza, R. & Kato, T. (2000). Features are also important: contributions of featural and configural processing to face recognition. *Psychological Science*, 11, 429-433.
- Chance, J. & Goldstein, A.G. (1976). Recognition of faces and verbal labels. *Bulletin of the Psychonomic Society*, 7, 384-386.
- Christie, D., Davies, G., Shepherd, J. & Ellis, H. (1981). Evaluating a new computer-based system for face recall. *Law and Human Behaviour*, 5, 209-218.
- Christie, D.F.M. & Ellis, H.D. (1981). Photofit constructions versus verbal descriptions of faces. *Journal of Applied Psychology*, 66, 358-363.
- Clifford, B.R. & Davies, G. (1989). Procedures for obtaining identification evidence. In D.C. Raskin (Ed.). *Psychological methods in criminal investigation and evidence* (pp. 47-95). New York: Springer Publishing Company, Inc.
- Cohen, M.E. & Nodine, C.F. (1978). Memory processes in facial recognition and recall. *Bulletin of the Psychonomic Society*, 12, 317-319.
- Comish, S.E. (1987). Recognition of facial stimuli following an intervening task involving the Identikit. *Journal of Applied Psychology*, 72, 488-491.
- Cutler, B.L. & Penrod, S.D. (1988). Context reinstatement and eyewitness identification. In G.M. Davies & D.M. Thomson (Eds.). *Memory in context: Context in memory*. Chichester: John Wiley & Sons Ltd.
- 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.
- Cutler, B.L., Stocklein, C.J. & Penrod, S.D. (1988). An empirical examination of a computerised facial composite production system. *Forensic Reports*, 1, 207-218.
- Da Costa, L.M. (1998). *Computer-guided facial reconstruction: an in-view and from-memory analysis of an eigenface-based system*. Unpublished honours thesis, University of Cape Town, South Africa.

- Davies, G. & Christie, D. (1982). Face recall: an examination of some factors limiting composite production accuracy. *Journal of Applied Psychology*, 67, 103-109.
- Davies, G. & Milne, A. (1985). Eyewitness composite production. A function of mental or physical reinstatement of context. *Criminal Justice and Behaviour*, 12, 209-220.
- Davies, G. & Oldman, H. (1999). The Impact of Character Attribution on Composite Production: A Real World Effect? *Current Psychology*, 18, 128-139.
- Davies, G. Ellis, H. & Shepherd, J. (1985, May 16). Wanted – faces that fit the bill. *New Scientist*, 106, 26-29.
- Davies, G., Ellis, H. & Christie, D. (1981). Exploring new strategies for facial recall. *Medical Science and Law*, 21, 137-145.
- Davies, G., Ellis, H. & Shepherd, J. (1978). Face identification: the influence of delay upon accuracy of Photofit construction. *Journal of Police Science and Administration*, 6, 35-42.
- Davies, G.M. & Valentine, T. (2007). Facial composites: forensic utility and psychological research. In R.C.L. Lindsay, D.F. Ross, J.D. Read & M. P. Toglia. *Handbook of eyewitness psychology. Volume 2: Memory for people* (pp. 59-83). Mahwah: Lawrence Erlbaum Associates.
- Davies, G.M. (1981). Face recall systems. In G. Davies, H. Ellis, & J. Shepherd (Eds.). *Perceiving and remembering faces* (pp.227-250). London: Academic Press.
- Davies, G.M. (1983). Forensic face recall: the role of visual and verbal information. In S.M.A. Lloyd-Bostock & B.R. Clifford (Eds.). *Evaluating Witness Evidence* (pp. 103-123). Chichester: John Wiley & Sons Ltd.
- Davies, G.M., Shepherd, J.W. & Ellis, H.D. (1978). Remembering faces: acknowledging our limitations. *Journal of the Forensic Science Society*, 18, 19-24.
- Davies, G.M., van der Willik, P. & Morrison, L.J. (2000). Facial composite production: a comparison of mechanical and computer-driven systems. *Journal of Applied Psychology*, 85, 119-124.
- Dodson, C.S., Johnson, M.K. & Schooler, J.W. (1997). The verbal overshadowing effect: why descriptions impair face recognition. *Memory and Cognition*, 25, 129-139.
- Ellis, H.D. (1975). Recognizing faces. *British Journal of Psychology*, 66, 409-426.

- Ellis, H.D. (1986). Face recall: a psychological perspective. *Human Learning*, 5, 186-196.
- Ellis, H.D., Davies, G.M. & Shepherd, J.W. (1978a). A critical examination of the Photofit system for recalling faces. *Ergonomics*, 21, 297-307.
- Ellis, H.D., Davies, G.M., & Shepherd, J.W. (1978b). Remembering pictures of real and 'unreal' faces: some practical and theoretical considerations. *British Journal of Psychology*, 69, 467-474.
- Ellis, H.D., Sheperd, J.W., & Davies, G.M. (1975). An investigation of the use of the Photofit technique for recalling faces. *British Journal of Psychology*, 66, 29-37.
- Ellis, H.D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, 8, 431-439.
- Farah, M.J., Wilson, K.D., Drain, M. & Tanaka, J.N. (1998). What is "special" about face perception? *Psychological Review*, 105, 482-498.
- Finger, K. & Pezdek, K. (1999). The effect of the cognitive interview on face identification accuracy: release from verbal overshadowing. *Journal of Applied Psychology*, 84, 340-348.
- Fisher, R.P, McCauley, M.R. & Geiselman, R.E. (1994). Improving eyewitness testimony with the cognitive interview. In D.F. Ross, J.D. Read & M.P. Toglia (Eds.). *Adult eyewitness testimony* (pp. 245-269). Cambridge: Cambridge University Press.
- Fisher, R.P., Geiselman, R.E. & Amador, M. (1989). Field test of the cognitive interview: enhancing the recollection of actual victims and witnesses of crime. *Journal of Applied Psychology*, 74, 722-727.
- Frowd, C.D. (2001). *EvoFIT: A Holistic, Evolutionary Facial Imaging System*. Unpublished PhD thesis, University of Stirling, UK. [<http://hdl.handle.net/1893/54>]
- Frowd, C.D., Bruce, V., McIntyre, A.H., Ross, D., Fields, S., Plenderleith, Y., & Hancock, P.J.B (2006). Implementing Holistic Dimensions for a Facial Composite System. *Journal of Multimedia*, 1, 42-51.
- Frowd, C.D., Bruce, V., Ness, H., Bowie, L., Paterson, J., Thomson-Bogner, C., McIntyre, A. & Hancock, P.J.B. (in press). Parallel approaches to composite production: interfaces that behave contrary to expectation. *Ergonomics*.

- Frowd, C.D., Bruce, V., Storås, K., Spick, P. & Hancock, P.J.B. (2006). *An evaluation of morphed composites constructed in a criminal investigation*. In proceedings of the 16th Conference of the European Association of Psychology and Law, pp. 59-66, London: IP-PA Publishing.
- Frowd, C.D., Carson, D., Ness, H., McQuiston-Surrett, D., Richardson, J., Baldwin, H. & Hancock, P. (2005). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology, 10*, 63–81.
- Frowd, C.D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S. & Hancock, P. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law, 11*, 33-52.
- Frowd, C.D., Hancock, P.J.B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perceptions, 1*, 19-39.
- Frowd, C.D., McQuiston-Surrett, D., Anandaciva, S., Hutton, C. & Hancock, P.J.B. (in press). An evaluation of US systems for facial composite production. *Ergonomics*.
- Frowd, C.D., McQuiston-Surrett, D., Kirkland, I. & Hancock, P. (2004). The process of facial composite production. In A. Czerederecka, T. Jaskiewicz-Obydzinska, R. Roesch & J. Wojcikiewicz (Eds.). *Forensic Psychology and Law* (pp. 140-152). Krakow: Institute of Forensic Research Publishers.
- Geiselman, R.E., Fisher, R.P., MacKinnon, D.P. & Holland, H.L. (1986). Enhancement of eyewitness memory with the cognitive interview. *The American Journal of Psychology, 99*, 385-401.
- Gibling, F. & Bennett, P. (1994). Artistic enhancement in the production of Photo-fit likenesses. *Psychology, Crime and Law, 1*, 93-100.
- Gibling, F. & Davies, G. (1988). Reinstatement of context following exposure to post-event information. *British Journal of Psychology, 79*, 129-141.
- Gillenson, M.L. & Chandrasekaran, B. (1975). A heuristic strategy for developing human facial images on a CRT. *Pattern Recognition, 7*, 187-196.
- Green, D.L. & Geiselman, R.E. (1989). Building composite facial images: effects of feature saliency and delay of construction. *Journal of Applied Psychology, 74*, 714-721.

- Haig, N.D. (1986). Investigating face recognition with an image processing computer. In H.D. Ellis, M.A. Jeeves, F. Newcombe & A. Young (Eds.). *Aspects of Face Processing* (pp. 410-425). Lancaster: Martinus Nijhoff Publishers.
- Hancock, P.J.B. & Frowd, C.D. (2001). Evolutionary generation of faces. In P.J. Bentley and D.W. Corne (Eds.). *Creative Evolutionary Systems*. London: Academic Press.
- Hancock, P.J.B. (2002). Improving recognition rates using $\frac{3}{4}$ views in ABM's PRO-fit 3.1 facial composite system. *ABM White Paper*, ABM United Kingdom Limited.
- Hancock, P.J.B., Bruce, V. & Burton, M.A. (1998). A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Research*, 38, 2277-2288.
- Hancock, P.J.B., Burton, A.M. & Bruce, V. (1996). Face processing: human perception and principal components analysis. *Memory and Cognition*, 24, 26-40.
- Hasel, L.E. & Wells, G.L. (2006). Catching the bad guy: morphing composite faces helps. *Law and Human Behaviour*. Retrieved November 15, 2006, from www.springerlink.com.
- Hines, D. & Braun, J.A. (1990). Order of feature recognition in familiar and unfamiliar faces. *Brain and Cognition*, 14, 165-184.
- Hole, G.J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, 23, 65-74.
- Homa, D., Haver, B. & Schwartz, T. (1976). Perceptibility of schematic face stimuli: evidence for a perceptual Gestalt. *Memory and Cognition*, 4, 176-185.
- Kerr, N.H. & Winograd, E. (1982). Effects of contextual elaboration on face recognition. *Memory and Cognition*, 10, 603-609.
- King, D. (1971). The use of Photo-fit 1970-1971 – a progress report. *Police Research Bulletin*, 18, 40-44.
- Kirby, M. & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterisation of human faces. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 12, 103-108.
- Kitson, T., Darnbrough, M. & Shields, E. (1978). Lets face it. *Police Research Bulletin*, 30, 7-13.

- Koehn, C.E. & Fisher, R.P. (1996). Constructing facial composites with the Mac-a-Mug Pro system. *Psychology, Crime and Law*, 3, 215-224.
- Koehn, C.E. (1995). *Theoretical and applied evaluations of facial composite systems*. Unpublished PhD thesis, Florida International University, Florida.
- Koehn, C.E., Fisher, R.P. & Cutler, B.L. (1999). Using cognitive interviewing to construct facial composites. In D. Canter & L. Alison (Eds.). *Psychology and criminal detection* (pp. 41-63). Brookfield: Dartmouth.
- Kovera, M.B., Penrod, S.D., Pappas, C. & Thill, D.L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology*, 82, 235-246.
- Krafka, C. & Penrod, S.D. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology*, 49, 58-69.
- Laughery, K., Rhodes, B. & Batten, G. (1981). Computer-guided recognition and retrieval of facial images. In G. Davies, H. Ellis, & J. Shepherd (Eds.). *Perceiving and remembering faces* (pp. 251-269). London: Academic Press.
- Laughery, K.R. & Fowler, R.H. (1980). Sketch artist and Identikit procedures for recalling faces. *Journal of Applied Psychology*, 65, 307-316.
- Laughery, K.R., Duval, C. & Wogalter, M.S. (1986). Dynamics of face recall. In H.D. Ellis, M.A. Jeeves, F. Newcombe & A. Young (Eds.). *Aspects of Face Processing* (pp.373-387). Dordrecht: Martinus Nijhoff Publishers.
- Leder, H. & Bruce, V. (2000). When inverted faces are recognized: the role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology*, 53A, 513-536.
- Light, L.L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 212-228.
- Luu, T.N. & Geiselman, R.E. (1993). Cognitive retrieval techniques and order of feature construction in the formation of composite facial images. *Journal of Police and Criminal Psychology*, 9, 34-39.
- Malpass, R.S. & Devine, P.G. (1981). Guided memory in eyewitness identification. *Journal of Applied Psychology*, 66, 343-350.

- Malpass, R.S. (2004). A lineup evaluation "do-it-yourself kit" for attorneys and law enforcement. <http://eyewitness.utep.edu/diy.html>
- Malpass, R.S., Tredoux, C.G. & McQuiston-Surrett, D. (2007). Lineup construction and lineup fairness. In R. Lindsay, D. F. Ross, J. D. Read, and M. P. Tolia (Eds.). *Handbook of Eyewitness Psychology, Vol. 2: Memory for People* (pp. 155-178). Mahwah: Lawrence Erlbaum Associates.
- Matthews, M.L. (1978). Discrimination of Identikit constructions of faces: evidence for a dual processing strategy. *Perception and Psychophysics*, 23, 153-161.
- Mauldin, M.A. & Laughery, K.R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology*, 66, 351-357.
- McKelvie, S.J. (1976). The role of eyes and mouth in the memory of a face. *American Journal of Psychology*, 89, 311-323.
- McKelvie, S.J. (1981). Sex differences in memory for faces. *Journal of Psychology*, 107, 109-125.
- McQuiston, D.E. & Malpass, R.S. (2000, March). *Use of facial composite systems in U.S. law enforcement agencies*. Poster presented at the meetings of the American Psychology-Law Society, New Orleans, Los Angeles.
- Meissner C.A., Brigham J.C. & Kelley C.M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory and Cognition*, 29, 176-186.
- Meissner, C.A. & Brigham, J.C. (2001a). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, 15, 603-616.
- Meissner, C.A., & Brigham, J.C. (2001b). Thirty years of investigating the own-race advantage in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, 7, 3-35.
- Osterburg, J.W. & Ward, R.H. (1997). *Criminal investigation: a method of reconstructing the past*. Cincinnati: Anderson Publishing.
- Prag, B. (2000). *Investigating the effectiveness of a computer-driven system for reconstructing faces*. Unpublished honours thesis, University of Cape Town, South Africa.
- Prag, B. (2005). *Testing featural and holistic composite technologies under different methods of face recall*. Unpublished master's thesis, University of Cape Town, South Africa.

- Rakover, S.S. (2002). Featural vs. configurational information in faces: a conceptual and empirical analysis. *British Journal of Psychology*, 93, 1-30.
- Rosenthal, Y. (1998). *A computerised face recall system using eigenfaces*. Unpublished master's thesis, University of Cape Town, South Africa.
- Rosenthal, Y., de Jager, G. & Greene, J. (1998). *A computerised face recall system using eigenfaces*. In Proceedings of the Eighth Annual South African Workshop on Pattern Recognition, pp. 53–57, Rhodes University, Grahamstown.
- Schooler, J.W. & Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology*, 22, 36-71.
- Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, 75, 221-242.
- Shapiro, P.N. & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, 100, 139-156.
- Shepherd, J., Ellis, H. & Davies, G. (1981). Studies of cue saliency. In G. Davies, H. Ellis, & J. Shepherd (Eds.). *Perceiving and remembering faces* (pp. 105-131). London: Academic Press.
- Shepherd, J.W. & Ellis, H.D. (1996) Face recall – methods and problems. In S.L. Sporer, R.S. Malpass & G. Koehnken (Eds.). *Psychological issues in eyewitness identification* (pp. 87-115). New Jersey: Lawrence Erlbaum.
- Shepherd, J.W., Davies, G.M. & Ellis, H.D. (1978). How best shall a face be described? In M.M. Gruneberg, P.E. Morris & R.N. Sykes (Eds.). *Practical aspects of memory* (pp. 278-285). London: Academic Press.
- Smith, S.M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human learning and memory*, 5, 460-471.
- Smith, S.M. (1984). A comparison of two techniques for reducing context-dependent forgetting. *Memory and Cognition*, 12, 477-482.
- Sullivan, T. (2004). *Guided Memory as an Aid in Reconstructing Faces Using the Face Composite System, ID*. Unpublished honours thesis, University of Cape Town, South Africa.

- Tanaka, J.W. & Farah, M.J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, 46A, 225-245.
- Tanaka, J.W. & Sengco, J.A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, 25, 583-592.
- Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception*, 9, 483-484.
- Tredoux, C., da Costa, L., Nunez, D. & Rosenthal, Y. (1999, July). *Face reconstruction using a configural, eigenface-based composite system*. Paper presented at SARMAC III, Boulder, Colorado.
- Tredoux, C., Nunez, D., Oxtoby, O. & Prag, B. (2006). An evaluation of ID: an eigenface based construction system. *South African Computer Journal*, 37, 90-97.
- Tredoux, C., Nunez, D., Oxtoby, O., Rosenthal, Y., da Costa, L. & Prag, B. (December, 2001). *Eface – a face composite system*. FIU presentation.
- Tredoux, C.G. (2001). *A Non-Verbal Approach To Face Reconstruction by Eyewitnesses*. NRF Technical report on CSD grant 15/1/3/16/0158.
- Turk, M. & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3, 71-86.
- Walker-Smith, G.J., Gale, A.G. & Findlay, J.M. (1977). Eye movement strategies involved in face perception. *Perception*, 6, 313-326.
- Wells, G.L. & Hryciw, B. (1984). Memory for faces: encoding and retrieval operations. *Memory and Cognition*, 12, 338-344.
- Winograd, E. & Rivers-Bulkeley, N.T. (1977). Effects of changing context on remembering faces. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 397-405.
- Wogalter, M.S. & Marwitz, D.B. (1991). Face composite construction: in-view and from-memory quality and improvement with practice. *Ergonomics*, 34, 459-468.
- Young, A.W., Hellawell, D. & Hay, D.C. (1987). Configurational information in face perception. *Perception*, 16, 747-759.
- Yuille, J.C. (1993). We must study forensic eyewitnesses to know about them. *American Psychologist*, 48, 572-573.

Appendix A

Photo protocol

Photographing faces: the Schaupp-Tredoux protocol

- 1 Physical-spatial preparation
 - a) Place **chair, camera and flash units** in the designated places. If they are already setup, check their position.
 - b) Set flash units to the following heights; Lower flash unit = 78 cm; Higher flash unit = 128 cm.
Refer to Drawing A.

- 2 Setup of flash heads
 - a) Check that switch on wall is on.
 - b) Power both flash units on, including modeling lights (to II, higher power setting).
 - c) Each flash unit to have flash intensity set to maximum.
 - d) Discharge each flash unit twice to dissipate pre-charge.
Refer to Drawing B.

- 3 Setup of camera
 - a) Mount camera on tripod, in portrait orientation; ensure camera is level and straight.
 - b) Connect flash sync lead to camera and to lower flash head.
 - c) Set mode to M(annual) on mode selection wheel.
 - e) Set image quality to large, smooth.
 - f) Set white balance option to color temperature (K) on small LCD screen.
 - g) Set color temperature to 5000 K (via menu).
 - h) Set aperture to 8.0, shutter speed to 60 (i.e. $f = 8.0$, $s = 1/60^{\text{th}}$).

- 4 Photo subjects
 - I Preliminaries**
 - a) Explain release form to subject and get signature.
 - b) Ask subject to remove jewellery, headgear, glasses. If they wear heavy make up, continue, but ask them at the end of the session if they can return later without make up.
 - c) Ask subject to sit down in chair.
 - d) Adjust height of tripod so the camera lens is at eye level of subject.
 - e) Adjust angles of the umbrellas and flash units to be 45 degrees to subject's line of sight.
(see Drawing A)

- f) Adjust angles of the flash units according to qualitative impression – lower unit umbrella must reflect downward, higher unit must reflect upward.
- g) Take photograph of MacBeth colour chart for first subject of the day/session. Subject to hold chart so that it is centred in frame, parallel plane to camera. Tell the subject to hold colour chart on the edges without touching any of the colour swats on the chart.

II Photographs

- h) Only one subject in the room at a time. Subject to sit down, upright, not against back of chair. In frontal view should fill 2/3 of frame horizontally, ½ of the frame vertically
- i) Tell subject that you are taking four shots
- j) Take casual shot; check for this and all other shots that results are acceptable i.e. no shut eyes, in frame, in focus (test focus with zoom view option in review mode of camera)
- k) Take frontal shot, but VERY IMPORTANTLY ensure
 - i) Vertical orientation angle to be checked (“lift chin”, “drop chin”; lead by example)
 - ii) Horizontal orientation angle to be checked (“swivel your head to the left”, lead by example)
 - iii) Pupil eye line angle to be checked (“tilt your head to the left/right” etc. lead by example)
 NOTE that these are dynamic and changes in one can undo changes in another! Check and re-check.
- l) Take ¾ shot. Ask subject to change seating position, not just head angle, and to look to back of umbrella. Adjust vertical (“drop chin”) and horizontal orientation (“swivel left”). A correctly composed shot will just show a hint of the far cheek and eye.
- m) Take 90 degree profile shot. Check vertical and horizontal angles
- n) Thank subject; say goodbyes











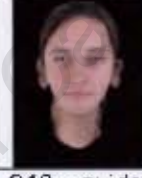


















4 Closure

















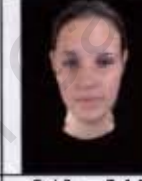








- a) Switch off camera
- b) Turn flash units off, power at wall off
- c) Take out sync leads from camera
- d) Dismount camera.
- e) Camera to be removed from venue, door to be locked
- f) Memory flash card to be given to Heike. Heike to download images, copy them to the Macintosh server, and to add them to the backup CD or DVD.





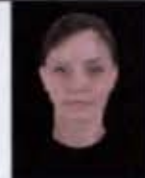




















Key to Taryn

Appendix B

Perpetrator photographs and the 76 composites constructed

PERPETRATOR ONE					
					
					
C1 – in view ID	C2 – in view ID	C3 – in view ID	C4 – in view FACES	C5 – in view FACES	C6 – in view FACES
					
C7 – in view FACES	C8 – in view FACES	C9 – guided ID	C10 – guided ID	C11 – guided ID	C12 – guided FACES
					
C13 – guided FACES	C14 – guided FACES	C15 – guided FACES	C16 – guided FACES	C17 – guided FACES	C18 – guided FACES
					
C19 – SAP ID	C20 – SAP ID	C21 – SAP FACES	C22 – SAP FACES	C23 – SAP FACES	C24 – SAP FACES
					
C25 – expert ID	C26 – expert FACES	C27 – morph ID	C28 – morph FACES		

PERPETRATOR TWO					
					
					
C1 – in view ID	C2 – in view ID	C3 – in view ID	C4 – in view ID	C5 – in view FACES	C6 – in view FACES
					
C7 – in view FACES	C8 – in view FACES	C9 – guided ID	C10 – guided ID	C11 – guided FACES	C12 – guided FACES
					
C13 – SAP ID	C14 – SAP ID	C15 – SAP ID	C16 – SAP ID	C17 – SAP FACES	C18 – SAP FACES
					
C19 – SAP ID	C20 – SAP FACES	C21 – expert ID	C22 – expert FACES	C23 – morph ID	C24 – morph FACES

PERPETRATOR THREE					
					
					
C1 – in view ID	C2 – in view ID	C3 – in view ID	C4 – in view ID	C5 – in view FACES	C6 – in view FACES
					
C7 – in view FACES	C8 – in view FACES	C9 – guided ID	C10 – guided ID	C11 – guided ID	C12 – guided ID
					
C13 – guided FACES	C14 – guided FACES	C15 – SAP ID	C16 – SAP ID	C17 – SAP ID	C18 – SAP ID
					
C19 – SAP FACES	C20 – SAP FACES	C21 – expert ID	C22 – expert FACES	C23 – morph ID	C24 – morph FACES

The four trial faces



Appendix C

South African Police interview used in Study 1

MAS/CAS	
<u>BESKRYWING VAN VERDAGTE/S</u>	<u>DESCRIPTION OF SUSPECT/S</u>
<p>A. RAS/RACE</p> <p>B. GESLAG/SEX</p> <p>C. OUDERDOM/AGE</p> <p>D. LENGTE/HEIGHT</p> <p>E. GELAATSKLEUR/COMPLECTION</p> <p>F. LIGGAAMSBOU/BUILD</p> <p>G. HARE/HAIR</p> <p>H. BAARD/BEARD</p> <p>I. SNORM/MOUSTACHE</p> <p>J. GESIG/FACE</p> <p>K. NEUS/NOSE</p> <p>L. LIPPE/LIPS</p> <p>M. Oë/EYES</p> <p>N. MOND/MOUTH</p> <p>O. KEN/CHIN</p> <p>P. TANDE/TEETH</p> <p>Q. VOORHOOF/FOREHEAD</p> <p>R. ORE/EARS</p>	
<p><u>EIENAARDIGHEDE VAN GELAAT OF LIGGAAM, TATOEËR- OF ANDER MERKE/ FACIAL OR BODILY PECULIARITIES, TATTOO OR OTHER MARKS</u></p>	
<p>KLEREDRAG/CLOTHING</p> <p>GEWOONTES/HABITS</p> <p>JUWELIERSWARE/JEWELLERY</p> <p>ANDER BESONDERHEDE (Opmerkings bv. Stem)</p> <p>OTHER PARTICULARS (Remarks on Voice)</p>	
<p>DATUM/DATE</p> <p>PLEK/PLACE</p>	
<p>HANDT. GETUIE/SIGNATURE WITNESS</p> <p>POLISIEBEAMPTTE/POLICE OFFICER</p>	

Appendix D

Lineups used for the evaluation tasks *

Lineups for matching task



Lineups for similarity rating task



* Note: each member varied randomly in placement for each task, as controlled for by a statistical randomisation programme.