Determinants of Estimated Face Composite Quality

Heike Christine Schmidt

Thesis submitted
for the degree of doctor of Philosophy (Ph.D.),
in the Department of Psychology,
University of Cape Town

Supervisor: Professor Colin G. Tredoux

February, 2010

Acknowledgements

This PhD thesis would not have been possible without the support of many. To them I dedicate this thesis.

To my parents Erika and Volker, who raised me to take on challenging tasks – and see them through.

To my supervisor Colin Tredoux, who nurtured my academic skills whenever I needed it – I could not have asked for more.

To the men and women of the Identikit-Unit and the LCRC Cape Town, in particular to Supt. Van der Westhuizen and Sen. Supt. April, who believed in me and my research enough to spend the time and effort required to make the field studies happen. To the well over 400 investigating officers who took time to supply the information. To the SAPS and Identikit Management in Pretoria and the Western Cape, who gave me permission to do the research, and who saw to it that I found open doors when I came knocking.

To the staff of the Department of Psychology, who accommodated me and my experimental studies and supported me in every aspect of my PhD. To my four interns Nonye, Julia, Nils and Katharina and my assistant Katherine, who helped me in the studies. To Dave, without whose help there would not have been an online-part of the experiment.

To my friends, old and new, far and near, who supported me in every stage of the way. To Carol, Kiki, Ridwana, Sarah and my wonderful Uli – without you all I would have never managed.

I would also like to thank the German Academic Exchange Service and the Postgraduate Funding Office at UCT for partial funding of this research.

Abstract

This thesis addresses the evaluation of an investigative tool commonly used by police forces around the world, namely a face composite or facial likeness. Face composites are constructed by witnesses to crimes, and are used to further the investigation. The process of constructing a composite involves a number of cognitive elements, all of which contribute to the final composite quality. This thesis examines elements of the construction process and assessment of the final composite quality in research and practice.

There are three main aspects to the empirical work reported in this thesis. The first, consisting of two experimental studies, investigates the possibility of reinstating context as a way of improving composite quality. The first of these studies explores featural and holistic influences of composite systems and interview methods on composite quality. Results of this study showed no superiority of either the featural or the holistic approach in consistently producing high quality composites. The second experimental study examined the influence of mental and physical context reinstatement on composite quality. The results of this study provided some evidence that context-reinstatement, although having a strong effect on the recall of crime details, does not have an equally strong overall effect on composite quality.

The second aspect of the work on which this thesis reports examines composite construction and use within the South African Police Service. Two field studies investigated police practice by examining police dockets, survey forms, and existing databases. Results showed that composite construction in real-life is in many aspects different from composite construction in laboratory experiments, and from composite construction as advocated by police policy.

The third and last aspect of the empirical work examined the measurement of composite quality itself. An experimental study compared the most common composite quality measurements, using the same stimulus material and composites. Results suggested that composite quality consists of multiple dimensions, and that these are related to the measurement method used. Specifically, several measurements provided evidence for the importance of a distinction between diagnosticity or accuracy, and willingness to make a choice, irrespective of the accuracy of that choice. The thesis concludes by arguing that composite quality can be regarded as a multidimensional concept and by advocating further investigation into this concept, which serves as the basis on which composites are judged.

In summary, this thesis has investigated cognitive factors influencing composite construction and composite quality measurement. It has found that clearer definitions of cognitive concepts related to these tasks are needed. In this respect, research on face composites is in want of clearer guidelines by which to judge the appropriateness of an experimental design for evaluating cognitive phenomena.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis addresses the evaluation of an investigative tool commonly used by police forces around the world, namely a face composite or facial likeness. A face composite is constructed by the witness or victim of a crime to aid the police in the investigation. As a composite is based on a witness's memory, research on composite construction is part of the domain of eyewitness testimony. Eyewitness testimony is a critical component in solving crime, and might in some investigations even be the only investigative aid available to help discovering the culprit (Deffenbacher, 2008; Wells & Olson, 2003). A composite might be used as investigative aid to narrow down the pool of possible suspects or produce additional information from informants or the public, thus directing the investigation towards a limited number of suspects that warrant further investigation.

The main idea behind composites is related to the old saying "a picture is worth more than a thousand words". In this case it expresses the belief that a composite, as a visual likeness of a face, can express the memory of a person for that face better than any verbal description can. This is based on the notion that the vocabulary to accurately describe a face and its features is mostly not readily available, and that a witness to a crime is not always able to accurately describe a face verbally, even though she might have a perfect memory of that same face. Moreover, constructing a visual likeness of a visual memory prevents distortions from the translation into a different modality. This, as well as all other topics raised here, will be discussed in more detail in subsequent chapters.

For the sake of expedience, witnesses and participant-witnesses will be referred to as female, suspects and targets will be referred to as male. In studies on composites, most targets and participant-witnesses are of the gender to which they have been assigned, and as such I have decided, for simplicity, to refer to a single gender. This is by no means a reflection of stereotyping different groups, but rather a simplification of the writing and reading process. I acknowledge that both genders can be present in both groups.

The process of constructing a composite involves a number of cognitive elements or steps which are the same for the witness in a police setting and the participant-witness in an experimental setting. A witness first has to recall the face on which the composite is to be based. She then has to convert this visual image into a verbal description in order to communicate this memory to the police officer serving as system operator. Depending on the composite system used, she then has to compare either features in isolation or begin by comparing a whole face to her memory of this face. Based on this comparison she then has to decide what has to be changed in order to achieve a closer similarity between composite and her memory for the face, or a better recognizability of the composite. In order to do this, she either has to identify the facial part that needs changing, or to describe this change in any other way so that the operator can implement it. For example, the witness could state that his nose was shorter, or that he looked older than the current composite. After this change has been implemented, she again has to compare composite to facial memory and decide on the changes necessary to improve the composite. This iterative process continues until she is satisfied that the composite quality is as good as it can be, or until the time allocated for the construction process is spent.

One consideration that requires greater discussion is the question of expected composite quality. The philosophy underlying the first mechanical systems, such as Photofit (Penry, 1971) and the Field Identification System (Laughery, Smith, & Yount, 1980), is to produce a type likeness or general facial type. However, in more recent times, systems have been developed that strive towards a more photo-realistic impression of the face, often using photographs as the basis for their library of features, some of them in colour (Clifford & Davies, 1989). The interested reader may try to imagine the eye colour of a close friend or next-of-kin, and consider how much less likely it is that a witness will remember the eye-colour of a person whom she has seen only once. This trend for more detail and photo-realism is based on the notion that memory for a briefly observed unfamiliar face is detailed enough to be retrieved without difficulty, an assumption that will be discussed in greater detail in Chapter 2.

Even though creating a composite from the memory of a face is a demanding cognitive task, Brace, Pike, Kemp, Turner, & Bennett (2006) have confidence in the usefulness of composites as investigative tools. It is worth considering that composites are but one tool in a police investigation, and that a single response from publishing the composite may assist this investigation by pointing towards a specific suspect.

Psychological research into composite construction serves two purposes: knowledge gain and implementing this knowledge into best-practice policy (Malpass et al., 2008). It seems that in this domain, as in other areas of eyewitness testimony, past inquiries were driven less by the quest for a theoretical framework in which findings can be placed, but rather by the need for practical implementation (Wells & Olson, 2003). However, besides the feeling of 'slight homelessness' (Malpass, 1993, p. 80) when designing a study and interpreting its results in terms of the cognitive processes involved, it has led to a multitude of different and sometimes contradictory assumptions about composite construction, as will be made apparent throughout this thesis. Chapter 2 will try to remedy this situation by exploring already existing cognitive theories and investigated phenomena in the area of eyewitness testimony, and their value for the domain of composite construction research.

Many researchers have applied themselves to the problem of how to improve composite quality. Here, it is interesting to see that in this field of research contradictory findings do exist without stirring much controversy. For example, Koehn and Fisher (1997) describe the software system Mac-a-Mug Pro as unsuitable and not useful for realistic police settings. Yet, Kovera, Penrod, Pappas, and Thill (1997) base their recommended use of composites in general, or rather their recommended caution in doing so, on their studies of the efficacy of this particular system, discarded by Koehn and Fisher as one that is unsuitable construction of good-quality composites. Several years prior to these publications, one of the authors of the latter studies (Cutler, Stocklein, & Penrod, 1988) concluded that Mag-a-Mug Pro was an effective software program that enables recognition of the target person from a composite. Based on these three publications, it is impossible to draw a conclusion as to the suitability of this software program, or its capability to construct recognisable, and therefore useful, composites. It is my opinion that such controversies in the field of composite research occur due to the multitude of experimental methods used, and the lack of comparability between different methodologies. This topic will be discussed further in Chapter 3.

One promising possibility for improving composite quality is the use of context-reinstatement techniques before and during the construction process. As stated earlier, the witness is repeatedly asked to access her memory for the face, and to compare this mental image to the image in front of her. The operator can guide her by relying on interview techniques helping the witness in the construction process. As discussed in Chapters 2 and 3, it was found that under certain conditions appropriate interview techniques can help the witness recall the face, leading to a better-quality composite. Studies described in Chapter 4 were aimed at increasing the knowledge of those conditions. Two studies were implemented that investigated different context-reinstatement techniques in an attempt to ascertain which aspect of context-reinstatement is most successful in improving the final composite.

A further topic addressed in this thesis concerns the mismatch between best-practice recommendations and knowledge about current practice. Although experimental studies of aspects of the construction process are extensive, real-world data is sparse. The systematic study of police practice has been implemented in other areas of eyewitness testimony. I consider it necessary to supplement laboratory research on composites in two important aspects: firstly, the comparison between ideal conditions found in the laboratory and conditions found in current practice can stimulate police policy to move towards more favourable conditions. Secondly, the same comparison can be used to determine whether past laboratory studies are actually meaningful for police practice, and to adapt future laboratory studies to a more forensically valid setting if needed. For example, research and practice might differ very profoundly regarding the time interval between the witness's exposure to the crime and their construction of the composite. Laboratory research largely implements composite construction without any delay, as will be discussed in Chapter 3. Although police policy sometimes stipulates a maximum delay of 2-3 days between crime encounter and composite construction, police practice might differ in that composites might be constructed after far longer periods of time. It would be useful to compare police practice, police policy and best-practice as stipulated by laboratory research, and to investigate the effect of those different conditions on the final outcome. Accordingly, Chapter 5 will present two studies investigating police practice in South Africa. In these studies I explore the conditions under which composites are constructed, as well as their utility.

In order to make recommendations for best practice, research must provide evidence that one experimental condition, such as a specific software system or interview method, is indeed superior to others. It must therefore replicate the finding that this experimental condition produces better-quality composites in a variety of situations or circumstances, with a variety of measurement techniques. Research on composites is especially rich in techniques measuring composite quality. As discussed in Chapter 3, experiments on composite construction employ at least two different quality measures, most of them arriving at different results for the different techniques. In order to make measurement techniques more comparable, Chapter 6 will present a study exploring all commonly used measurements and their relationships to one another.

In response to Malpass's (1993) earlier statement of "homelessness", I would like to contribute a brick to the groundwork for a stable home on which research into composite construction can be built. It is up to more elaborate pieces of work by more senior researchers to build a permanent home. Through the evaluation of both cognitive theory as well as practical research I will conclude this thesis with a discussion regarding composite construction in Chapter 7.

There is clear need for improvement in existing research as well as a need for a more fundamental analysis of the domain of recalling a face for composite construction (Clifford & Davies, 1989). In this thesis I will attempt an analysis of existing research and present studies on psychological and methodological aspects of composite construction.

Chapter 2

Theoretical considerations of face composite construction

Many studies in the field of face composite construction have targeted specific practical issues, such as which composite system produces the best possible likeness of a depicted face. Research motivated by the desire to advance the theoretical understanding of face processing and face retrieval within the field of face composite construction has, however, been addressed to a more limited degree. Cognitive aspects of face processing have been studied in other contexts, and serve composite research as an explanation of different phenomena. What is missing is an attempt to find a theoretical foundation of face retrieval for composite construction. On the one hand, this enterprise is indeed very difficult because cognitive processes involved in composite construction often straddle two distinct phenomena, such as recall and recognition, or holistic and feature processing. On the other hand, it seems that purely practically motivated research runs the risk of being aimless or uncoordinated, leading to singular and non-replicated forays into the particular field of interest. In contrast, a theory-driven approach, asking questions about the underlying cognitive processes involved in composite construction, can provide a framework that can summarize past research and guide future research towards a more satisfying conclusion. It would be presumptuous to believe that a single thesis can develop such a framework. I will not offer a new theory or even modify an existing theory. But what this thesis will attempt to do is question and elaborate on fundamental processes underlying composite construction, which will be done in this chapter. Also, I will highlight some of the practical consequences that emanate from these theoretical considerations. This chapter will focus on a deeper theoretical understanding of composite construction, whereas the next chapter will focus on more methodological issues pertinent to research in this field.

The present research is carried out within the field of memory (i.e., face processing) and social interaction (i.e., context-reinstatement interview). Both fields have a vast and diverse theoretical and methodological tradition, which is only in part relevant to this piece of work. My focus is centred on concepts relevant to the presented research, with the passing acknowledgement of more detailed reviews of the various topics. Within the field of eyewitness testimony, many aspects of retrieval are defined as, for example, recognition and recall, holistic and configural, recognition and familiarity. All these concepts are related to the retrieval of a face for composite construction. Therefore this chapter will start by trying to specify why faces are a special category of stimuli, defining face composites, how composites came to be part of law enforcement, and how the above mentioned cognitive concepts help in understanding problems and challenges in composite construction. Thereafter, I will focus on concepts relating the memory for faces to memory of additional information, and how this network of encoded information can aid in retrieving the memory for a face.

2.1 Basic Concepts

What is a Face?

“Faces are arguably the most important biological and social objects in our environment” (Ellis, 1990, p. 144).

The emphasis on faces as being an important aspect of our life has been reiterated by many scientists, and the extent of research on cognitive and social aspects of face processing in humans reflects this importance. Being able to observe and remember a face is a matter of habit and practice. It is not a given for everybody even though everybody might see faces on a daily basis – one has to develop this skill (Hinkle, 1990). Faces are very similar in general structure, but people are very attentive to the subtle differences between them and can differentiate between faces on the basis of these minor differences. Indeed, Barry, Johnston, and Scanlan (1998) state that faces are probably the most homogenous set of stimuli we encounter. Even though almost all faces have the same structure, that is, two eyes, a mouth and a nose, all positioned within an oval that is relatively invariant, we are able to recognize many thousands of faces. Faces are not only the most reliable cue to distinguish between individuals, but also aid in gathering social information such as moods, feelings, attributed characteristics, age and gender (Young & Bruce, 1991). Regarding one of the properties of face memory, retention of facial memory over a period of time, Deffenbacher, Bornstein, McGorty, and Penrod (2008) found that memory for faces does not follow the same forgetting curve as common English words or pictures of common objects, but consolidates at a more rapid rate. Their analysis of past studies shows that memory for once-seen faces declines more slowly. This finding supports the notion of faces being important bearers of information, more so even than any other category of objects. Faces enable us, for example, to distinguish friend from foe, avoid an angry person when one sees rage in his/her face, and enter into a pleasant conversation upon seeing the other person smile. Faces provide a rich source of socially relevant information, giving us an evolutionary advantage if we can discriminate between them.

Regarding the perception of a face, the first step of this information-gathering process is to detect a face within an image, regardless of whose it is. This process has attracted the attention of information-technologists who strive to develop automatic face detection and recognition systems that can match the processing capacity of the human brain (Ellis & Shepherd, 1992). It might be possible to approximate human face perception by looking at advances in technology-based face perception. One way to determine whether a specific image contains a face is template matching: a computer-modelled template is swept across the image, and any region that matches the template is scored as a face (Tsao & Livingstone, 2008). Another approach is the use of a sequence of increasingly complex filters for feature detection (Viola & Jones, 2004). With their computer model they show that the presence of a face can be ruled out with two very simple filters, thus avoiding finer-scaled filtering of uninformative regions of the image. This cascade-approach with only two filters of rectangular dark-and-light regions used during initial encoding of a face (see Figure 2.1.) proved to be just as accurate as the template-single-step approach, only 10 times faster.

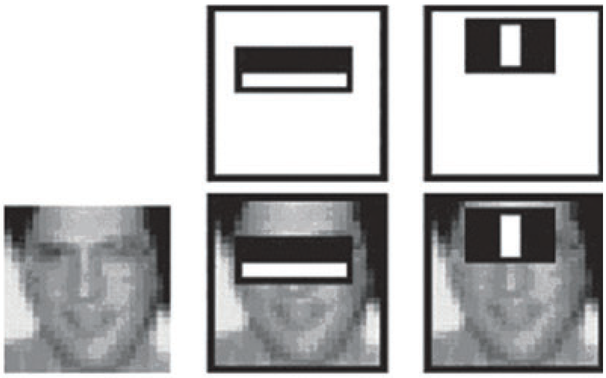

Figure 2.1. Sequential Feature Detectors for Face Recognition (after Tsao & Livingstone, 2008, p. 414)

Both algorithms use feature-detectors focussing on specific facial features. Yet they also perform holistic detection in that they detect faces as properly arranged wholes. Once a shape has been classified as a face, it has to be identified, a process which will be discussed later in this chapter. As can be seen from these examples, the detection of a face is not an elusive cognitive process, but the human ability can be modelled fairly successfully electronically.

Faces represent a “special” class of objects, but it is as yet uncertain if the cognitive systems involved in face processing are distinctly different from those used by other classes of objects, or if faces are special in degree instead of in kind. Evidence that face recognition is different comes from studies in the fields of neuropsychology, animal research, infants and face inversion. It is not clear, however, that any of the studies in any of these fields can provide conclusive evidence that faces are a class of objects that is processed differently from any other class of objects. For example, Lund (2001) concluded from studies in neuropsychology regarding prosopagnosic patients that face recognition is a distinctly different cognitive process. Conversely, Ellis and Young (1990) point out that many prosopagnosic patients have problems identifying other objects as well as faces, thus impairment may not be limited to face retrieval.

Tsao and Livingstone (2008) consider face identification as not being different to other forms of object recognition. Object discrimination in a class of objects sharing a common basic configuration is certainly demanding and requires particular sensitivity to differentiate, but is by no means special to the class of faces. We can differentiate between hundreds of bottles, at least pairwise, and we can recognize thousands of words at a glance, both requiring expert detection of configural patterns of a class of stimuli. Mechanisms of face processing are optimized due to practice and necessity, but are argued to be similar to other object processing (Casey & Newell, 2007; Diamond & Carey, 1986; Gauthier & Tarr, 1997). Diamond and Carey developed a "face expertise theory", which states that faces are not a special group of stimuli per se, but that expertise for this category is very developed in relation to other classes of stimuli. Supporting this theory, neuroimaging studies show that certain occipitotemporal regions react predominantly, but not exclusively, to faces (Haxby, Hoffman, & Gobbini, 2000). Haxby et al.'s interpretation of these findings is that if neurons that respond exclusively to faces exist, they are intermixed with neurons that respond to other objects. An alternative hypothesis by Gauthier, Behrmann, and Tarr (1999) and Gauthier, Tarr, Anderson, Skudlarski, and Gore (1999) states that areas responding mainly to faces must be specialized for visual expertise. Following this thought, any object that is perceived as an individual, be it a specific bird identified as a male zebra finch or a human as neighbour Thabo, will trigger a response in these brain regions. Since we are all experts at recognizing faces, these regions will be consistently activated in all subjects. Therefore, if faces prove to be processed in the same way as other objects, experimental findings and theoretical considerations can be implemented to research face composite construction.

What is a Face Composite?

Throughout the literature on face composites there are many different definitions of face composites and composite quality. These can be classified into three categories: ideal definitions, relating to memory processes; methodological definitions, relating to experimental designs; and applied definitions, relating composites to their use in police investigations. Many expressions are used in composite literature: face composite, facial likeness, and pictoral representation. In this thesis 'face composite', 'composite' and 'visual likeness' are used interchangeably to describe the visual representation of a target face, adhering to a more methodological than ideal definition.

If our memory was ideal, a composite would arise solely out of the memory for the face, and would be a "visual representation of a perpetrator's face" (Frowd et al., 2005a, p. 64). Following this idea, it would be possible, with the help of an ideal composite system, to construct a perfect likeness of a specific face. There are nearly seven billion faces on earth today (US Census Bureau, 2010), and every one of them is different from all the others (Hinkle, 1990). With the help of a composite system containing every possible feature, a witness could reconstruct this one face. However, neither are the systems currently in use perfect, nor is our memory (Wright, Memon, Skagerberg, & Gabbert, 2009). Our memory can be seen as a fragile entity, an associative and selective process of storing incoming information, and interpreting this stored information in the process of reassembly when it is called for. On the other hand, our memory for faces seems to be a rather uniquely capable skill, and enables us to recognize people we have not seen in 40 years, or recall a face we have seen only once.

Davies and Valentine offer a careful definition of face composites: "Police composites are impressions of a suspect's facial appearance derived from a witness description" (2007, p. 59). An equally careful definition is offered by Gibson, Solomon, Maylin, and Clark: "A facial composite system allows the expression of the facial appearance retained in the witness' memory" (2009, p. 156). This impression is a "pictoral representation" (Davies, Ellis, & Shepherd, 1978a) and should contain a "pictoral likeness" (Koehn, Fisher, & Cutler, 1999), although it may only be an imperfect likeness. Penry (1971) and Davies and Valentine use the term "type-likeness", placing emphasis on the fact that because a memory process is not ideal as stressed above, and because composite systems are not able to depict every possible nuance of a face, a composite will most likely not depict one specific individual that can be found within this population of seven billion human beings. A composite's value is in depicting a face type rather than a particular face. Therefore, even if composite systems do become better and more effective in depicting faces, as George, Gibson, Maylin, and Solomon argue is necessary, a "near photo-realistic facial composite" (2008, p. 1485) would be very nice to have, but is not a realistically likely outcome of face composite work. Also, research on caricatures suggests that diverting from photo-realistic likeness towards an exaggeration of key facial features may lead to a better recognizability than a veridical portrait (e.g., Lee et al., 2000).

A misperception seems to be that there is a correlation between a "good" composite and an accurate composite with a high resemblance to the suspect. A composite may look like a well-proportioned, realistic face, but this does not mean it contains a high resemblance to the target face. The outcome does not immediately imply that the process of construction was successful, that the witness's memory for the face was strong, and that the composite can be useful in finding the suspect in a police investigation (Taylor, 2001). The measurement of composite quality and/or usefulness will be discussed further in Chapter 3.

In most experiments published to date, a participant-witness is asked to inspect an unfamiliar face and afterwards to produce a face likeness with the help of a trained operator or on her own. Alternative experimental scenarios entail unintentional encoding of the face, that is to say, participant-witnesses are involved in a mock crime, a test situation or another situation in which they have to focus their attention on a target face. Generally, composites in experiments are constructed from a face seen only once before, leading Wells and Hryciw (1984) to label them "memory-for-face tests". An exception is the so-called in-view composite, usually constructed as a comparison for experimental conditions, or as a best-likeness option to assess the capability of the system.

Methodologically, a face composite is defined as the image of a face, designed by a participant-witness who saw the target face the composite is supposed to display, constructed with the help of either a sketch artist, a technician or on her own, on the basis of pencil-and-paper (or the equivalent computer screen) or a software program. In this thesis, participants in experiments, required to construct a composite after being exposed to a face, will be called participant-witnesses, constructing a composite from a target or target face. In a police environment these would be witnesses constructing a composite of the suspect or the suspect's face.

In police practice, face composites are seen by the general public in newspapers and on television programs, coupled with a plea to get in touch with the police should anyone know someone, or maybe have seen someone, bearing some resemblance to the published composite. Frowd et al. (2005a, 2005b) and Penry (1971) describe the goal of this publication as being a recognition of the face depicted in the composite as someone a member of the general public knows. Regarding their actual use in the United Kingdom, Kitson, Darnbrough, and Shields (1978) found that only 10% of composites constructed by the police services are released to the media, the remaining 90% being used for internal police enquiries.

As a screening tool, a composite can point to a suspect who might usefully be further investigated, or point away from an innocently investigated suspect who can be eliminated from further investigation (Ellis, 1986; Penry, 1971).

Police do not necessarily seek a 'pinpoint likeness' of the suspect, but rather seek to include a subset of suspects or, equally important, exclude suspects not resembling the composite (Davies & Valentine, 2006). For example, if a composite shows a rather slim-looking face, the chubby-faced, overweight suspect might not be eligible to be the focus of a further investigation, but police might rather focus on those suspects that are lean and slender. In other words, in an investigation a composite can support or reject the hypothesis that a certain suspect or a number of suspects are the culprit(s). If the composite resembles one of the suspects, the witness might be asked to identify the suspect in a line-up, and the investigating officer will search for further circumstantial and direct evidence (Wells et al., 1998).

History of Face Composites and Current Practice

The earliest technique of composite construction is the use of a police artist, who works in conjunction with the witness to construct a facial sketch. The construction of composites by sketch artists seems on the one hand still superior to modern composite systems (Frowd et al., 2005a), but is on the other hand not adhering to any standards regarding construction (e.g., use of photographs as reference material) or end-result (Cormack, 1979; Domingo, 1984; Taylor, 2001). Therefore the way sketch composites are constructed is dependent on the single artist, and no general conclusion can be drawn that an artist's sketch is always superior to a software-constructed composite. In the US there are more than 3,500 sheriff's departments (Davies & Valentine, 2006), but only 19 full-time artists (Poole, 2004). Since composite construction by hand requires a very specific set of skills, it is considered a "dying art", and most face composites are constructed with either mechanical or computerized systems (Davies & Valentine, 2006).

During the second half of the 20th century alternative approaches to the sketch artist were developed. These typically involve assembling pre-defined facial features into a whole. An advantage of this new mechanical technique is that composites can be produced by people not trained in portraiture. Two widely used systems, which became popular in the 1970s (Frowd et al., 2005a) are Photofit and Identikit[1], the former having recently been renamed Profit

[1]For Photofit and Identikit alternative spellings are sometimes used in this thesis and in the literature (Photo-Fit and Identi-Kit), referring to the same program.

(Schmidt & Frowd, 2006). Penry (1971) developed a hand-assembled facial composite system, Photofit, based on photographed features in picture card form, and implemented in Britain in the 1970s. While producing a composite, elements have to be fitted together like a jigsaw puzzle, with some vertical adjustment being possible. Photographs are also used as a basis for a catalogue compiled by the Federal Bureau of Investigation (Clifford & Davies, 1989). Here, witnesses have to choose a whole photograph for each feature, with the officer compiling the final composite. Another system, using drawings of facial features copied onto transparencies, was developed by a police officer in the US, and called Identikit in 1959 (Davies & Valentine, 2006). Here, the witness has to describe each feature, and the system operator picks the most appropriate transparency. The witness can then change and adjust each feature until she is satisfied with the result. Both Identikit and Photofit were marketed and used in more than 20 different countries, revised and extended over the years.

The crucial and critical assumption underlying these new feature-based composite kits was made explicit by Penry (1971), who considered the witness able to decompose a face into its distinct features, such as the eyes, nose and mouth. This assumption, which is the explicit base for these kits, sees a face as "simply the sum of its parts" (Bruce, 1988, p. 37), and the witness as able to add or subtract each part independently.

A third generation of composite construction methods is in part based on the same principle of feature-selection. The computer-based systems enabled an increase in the number of features available, the forms of alterations possible, the choice of a starting face, and the way in which features are displayed. Some of these systems currently in use start with a face specified by a prior description of the target face. E-Fit (Ellis & Shepherd, 1992) and Pro-Fit (Frowd et al., 2005a, 2005b) reduce the set of features initially available to the witness by obtaining a description of specific features prior to starting with the visual task of composite construction.

There are a number of reasons why one can expect computer-based systems such as Pro-Fit to be superior and produce better composites. For one, they do not require the witness to choose features in isolation from a catalogue of feature sets. Choosing features outside the context of the face is a purely featural task, and might be more difficult and more error-prone than relying on additional cognitive processes (Ellis & Shepherd, 1992). Also, pre-choosing fewer features from a description given prior to the construction reduces the search task for the best-fitting feature. On the other hand, one can argue that this limited range of features might not contain the most appropriate feature since the set was chosen on the basis of a verbal description. However, it must be considered why composites were first used in police investigations. Constructing a visual likeness seemed preferable to eliciting a verbal description of the face because witnesses, like most people, are not able to describe a face well enough (Hinkle, 1990; Penry, 1971). Thus the idea that a set of features, chosen from a description given by a witness untrained in face description and facial feature composition contains the most appropriate feature might be misleading. However, an indisputable advantage of computer-based systems lies in the adjustability of the configuration of selected features: features can be changed in size as well as moved in any direction.

The latest, fourth, generation of face composite systems uses genetic algorithms like principal-component analysis to generate a selection of faces from which the witness has to choose the best likeness. More specifically, Sirovich and Kirby (1987) developed the eigenface algorithm replicating human face processing on the computer. Subsequently developed computer algorithms based on principal-component analysis even outperformed human face identification in several tests developed for the 2006 ‘face recognition vendor (O'Toole, in press). Eigenface algorithms later became the basis for this new generation of composite systems. Three systems are currently tested in experiments, with some having been used by the police. ID (Tredoux, Rosenthal, Nunez, & da Costa , 1999), Evo-fit (Frowd et al., 2005a, 2005b; Hancock, 2000) and Eigenfit or EFit (Gibson, Solomon, & Pallares Bejarano, 2003) ask witnesses to choose a best-likeness from a set of computer-generated faces. Depending on their choice, a new subset of faces is generated, assuming that each new subset is closer to the witness's memory of the suspect's face than the one before. As a system already marketed, EFit-V presents the witness with a 3x3 array of randomly generated faces from a subsample, for example White males. Touching up on the face later, one can, for example apply a “wrinkle map” (Gibson et al., 2009), increasing the perceived age of the composite face.

For more than 50 years police forces have been using visual and later software aids to enable witnesses to form a face composite from their memory of the suspect's face. A review of this topic can be found in Davies and Valentine (2006), and a discussion on studies evaluating composite systems and adjacent factors in Chapter 3. None of those systems have until now shown a stark superiority in depicting good quality face composites (Frowd et al., 2005a, 2005b). Yet, if a witness should have an unlimited number of features, positions and expressions, the latest generation of systems based on genetic algorithms shows the most promise. Whether a future system will rely on the more holistic recognition-based cognitive process, or on the more feature-based process has still to be seen.

Recall or Recognition

Different composite systems place different cognitive demands on the witness. In the second generation of systems, as described by Davies and Valentine (2006), witnesses look through a catalogue of facial features and have to recognize the right, or most appropriate, feature. To reduce the number of features presented, some systems require a recall of facial features prior to construction. Other systems are based on the notion that whole face recognition is a stronger process than recall or in-parts-recognition, and that systems should be based on this process (e.g., Gibson et al., 2009). Thus, there are two types of processes involved in the task of composite construction: recall and recognition (e.g., Davies, 1983; Frowd, Bruce, Smith, & Hancock, 2008), usually considered distinctly separate processes (e.g., Sporer, 1989; Woodhead & Baddeley, 1981). In order to evaluate the merits of the different systems, I will evaluate the cognitive processes of recall and recognition in regards to composite construction.

Recall is defined as the retrieval of information without an external reference. It is assumed to be more difficult to describe a face than to recognize it, since it appears more effortful and more rapidly decaying (e.g., Bruce, 1982; Burton, Wilson, Cowan, & Bruce, 1999; Davies, 1983). Faces are difficult to describe, and maybe they ought not to be described because most people do not seem to have an adequate vocabulary to describe a face. In addition, if they describe it incorrectly, they might stand by their description rather than their visual memory.

The act of describing a face can, under certain conditions, interfere with the recognition of the face, a process that is sometimes referred to as the "verbal-overshadowing effect" (e.g., Dodson, Jonhnson, & Schooler, 1997; Meissner

& Brigham, 2001a; Schooler & Engstler-Schooler, 1990). Meissner, Brigham, and Kelley (2001) found that only when participants were forced to generate an elaborate description of the target face that later face recognition was impaired. Participants who received a warning to generate only correct descriptors benefited in their later recognition task. Meissner et al. concluded that manipulating retrieval processes of a face exerts a strong influence over face description and later face identification. Nevertheless, diminished recall ability through verbal overshadowing is not inevitable when describing a face under laboratory conditions. On the contrary, witnesses seem to be only susceptible under certain experimental conditions (for a meta-analysis see Meissner & Brigham, 2001a). Recalling facial features is, as will be described in Chapter 3, part of the interviewing process preceding composite construction in the British police. A potential verbal overshadowing effect, or the opposite, a 'verbal facilitation effect' (Meissner, Sporer, & Susa, 2008) might influence memory for the face, and subsequent retrieval during composite construction.

Recognition is a cognitive process that allows people to realize that what they perceive at present was perceived before. In other words, one realizes that a face was seen, or a song heard before. Rakover and Cahlon (2001) consider face recognition to be an ability that happens quite effortlessly and which may not require complex computation. Instead, they argue that it could be as simple and automatic as making footprints in the sand. Even though, intuitively, recognition seems to be the stronger cognitive retrieval process, it is certainly not without error. As many as one in five identifications might be faulty, meaning that a witness identifies an innocent bystander in the line-up (Valentine, Pickering, & Darling, 2003; Wright & McDaid, 1996). Recognition of a once-seen face is dependent on several elaborately defined variables, classified by Wells (1978) as estimator variables, aspects of the situation outside the control of the criminal justice system, and system variables, those aspects of the situation that are under the control of the criminal justice system. Both kinds of variables seem to have an effect on the ability of the witness to retrieve the suspect's face.

Recognition, in contrast to familiarity, sees the identification of a face as someone specific, often referring to a specific point in time when this person was encountered, and not as someone whose face seems to be more familiar than other faces (Bruce, 1988). In contrast to this, one sometimes has the feeling of familiarity, of having seen a face before, but not being able to consciously retrieve where and when this face was seen. Face recognition in a forensic setting usually requires more than recollection of the face itself, meaning that the person has to be identified as the robber seen at the bank, or the murderer seen at the home, or, more generally, the perpetrator seen at the crime. This implies a semantic level of representation including the information not only about the person, but also about time and place of the perception. A sense of familiarity alone is not viewed as sufficient to claim face recognition.

According to dual-process theory, two processes can bring about a feeling of recognition: recollection and familiarity (Yonelinas, 2002). Recollection is specified as the retrieval of episodic memory, a cognitive process strived for in eyewitness research. I might encounter a woman at a restaurant and recognize her as Angela whom I met on the train the day before, simply by recollecting exactly when it was that I saw her. Single-process signal detection theories suggest a single process of familiarity and that recognition is defined as familiarity above a certain threshold, with no additional cognitive process being postulated to explain recognition (e.g., Dunn, 2004). Using the same example as before, I might recognize Angela as familiar, without being able to identify when or where I saw her before. Discussions on the merits of each approach can be found elsewhere (Diane, Reder, Arndt, & Park, 2006; Jones & Bartlett, 2009; Yonelinas, 2002). The importance to eyewitness testimony is the differentiation between the feeling of familiarity and the actual recollection of the episode in one's life in which the specific person was encountered.

Adding to Ellis's (1984) three forms of recognition "tests" used in eyewitness practice, the latest generation of composite systems might be a fourth form. This new generation of eigenface-based software systems draws heavily on the process of recognition of whole faces (e.g., Frowd, Hancock, & Carson, 2004; Tredoux, 2001), as do all other face recognition tasks. The witness repeatedly has to choose the best likeness, based on his memory of the suspect, from a range of faces provided. Although some researchers have high expectations for these systems due to their reliance on a more effortless and automated cognitive process, these systems have yet to prove their superiority over other composite systems (Davies & Valentine, 2006).

Holistic or Featural Face Retrieval

Two types of face processing are usually mentioned in eyewitness literature: holistic processing and featural processing. Featural processing focuses on single facial features while the face is processed, manipulated in experiments either at encoding or at retrieval stage (Schwarzer, 2000). However, non-featural face processing can be defined in two different ways. Configural processing refers to the spatial relationships between features, a process without which features could not be placed within a face (Bruce, 1988; Diamond & Carey, 1986). To recognize a face, both featural and configural information is needed (O'Toole, 2005). By contrast, holistic face processing often refers to a face representation that exists as a whole, without being able to add up individual features and configurations to the same complete representation of a face (Farah, Wilson, Drain, & Tanaka, 1998). Holistic processing is characterized by the integrity of the face rather than its analysis and disintegration (Wells & Hryciw, 1984).

The 'holistic face processing hypothesis' (Hancock, 2000) assumes that faces are predominantly processed as a whole, much more so than other classes of stimuli, and that an analysis of features and configurations does not display all information present in a face. Tanaka and Farah (1993) suggest that upright face patterns cannot be broken down into parts at all, equating configural with holistic dimensions. Allowing three distinct facial dimensions, Cooper and Wojan (2000) suggest that face processing includes the processing of featural and configural information as well some global features or 'metric' variations. Exploring these distinctions empirically is difficult, since a feature change tends to result in a change in configuration, the holistic nature of the face as well as any global feature that might exist. However, Leder and Bruce (2000) undertook the task of separating configural and holistic face processing. In their experiment, faces that were recognizable by only showing combinations of features (brightness or colours) showed no effect when inverted. However, when the configuration of features was changed between different face identities, inverting the face showed a reduction in recognition performance. This result suggests that the inversion effect occurs, at least in part, due to a disruption of configural rather than holistic processing, ruling out the notion that only holistic face processing enables face recognition.

The rationale behind mechanical and computer-based featural systems assumes that a witness can dissect the memory of a face into individual features and compare these features to the features present in the composite kit. Research into face processing nevertheless suggests that faces are not only encoded as a set of features, but rather as a more general, undivided, representation in which feature information is incorporated into an impression of the face as a whole (Rakover, 2002; Tanaka & Farah, 1993, 2003; see review by Mauer, Le Grand, & Mondloch, 2002). It can, however, be argued that for currently available systems, there are no clear differentiations between featural and configural processes, as most of them enable the choice and alterations of features within the context of a face. It is, for example, possible to make purely configural changes to a face by keeping all of the features and simply moving them slightly. It is however less easy to make a purely featural change since changing a feature will change the configuration of a face (Bruce, 1988). However, apart from the latest generation of composite systems, witnesses constructing a composite have to analyse featural information and the configuration of those visible features.

Selecting features within the context of a face rather than in isolation aids in retaining configural information (Frowd et al., 2005a, Tanaka & Farah, 1993; Tanaka & Sengco, 1997), but still requires a certain dissection of the face. If faces are processed predominantly as holistic impressions, none but the latest generation of composite systems can ensure a "transfer appropriate processing" (Meissner et al., 2008) during retrieval of the face. If, however, faces can be processed as a combination of configural and featural information, composite systems should add configural aids as much as possible, thereby potentially keeping a more holistic than dissected impression of the face. Aiding configural processing as well as reducing the number of changes the witness has to make, it is for example possible to construct an initial full composite instead of having the witness choose features in isolation. Koehn et al. (1999) suggest that before the witness starts constructing a composite, a complete 'starting' face composite should be constructed by the computer from an initial description in order to adhere to a more holistic processing. Features can be chosen either from the description, or, if the feature isn't described, a middle or modal value can be put in as default. Any addition or change thereafter should be done within the context of the face.

Manipulating naturally occurring face processing can be done in all stages of face processing. However, only influences during encoding and retrieval are relevant for this thesis. During the initial perception of the suspect, witnesses cannot be influenced by police forces, but participant-witnesses in experiments can. Berman and Cutler (1998) and Wells and Hryciw (1984) believe that everyday facial perception is holistic in nature, but that it can be changed into featural processing when participants or witnesses are instructed to do so. In a study by Bower and Karlin (1974), instructions to judge the global impression of a face during encoding, for example the person's honesty or likeability, thus making an inferential judgement of the person seen, enhanced subsequent recognition performance. Wells and Hryciw (1984) and Wells and Turtle (1988) found that while character judgement improved recognition accuracy, featural judgement improved the accuracy of verbal description. They argue that featural judgement during encoding fosters featural analysis of the face. Conversely, judgements on more global impressions like personality and intelligence fosters the processing of more holistic dimensions, which might be beneficial for later face recognition. Likewise, in an experiment by Davies and Little (1990), a sketch artist who drew outlines of groups of features before specifying each feature constructed better quality composites when participant-witnesses were asked to encode the face using a holistic instead of a featural strategy. Koehn et al. (1999) propose a matching strategy between encoding and retrieval processes. A face that has been encoded holistically can only achieve a limited quality when retrieval is governed by a feature-based composite production system.

Witnesses in real life will in all likelihood not consider later face composite construction while witnessing a crime, and will therefore not try to memorize single features of the face (Koehn et al., 1999). Thus a witness' mental representation of the face is bound to be more holistic than featural. In an experimental setting, however, participant-witnesses are often bound to a more intentional learning strategy. Frowd et al. (2007a) suggest that when told that they will later have to construct a composite, witnesses will be more prone to carry out a featural encoding, which benefits feature-based quality-assessment including composite construction. A mock event with an unsuspecting participant-witness might be the only way in which unintentional learning of the target face can be ensured in an experiment.

Unfamiliar Face Processing

An influential model of face recognition, 'the interactive activation and competition model' (Bruce & Young, 1986; Burton et al., 1999; Hay & Young, 1982;

Young & Bruce, 1991; Young & Burton, 1999) focuses on the recognition of familiar faces, a process which is only marginally interesting for the retrieval of unfamiliar faces for composite construction. Processing unfamiliar faces is, although cognitively harder, the more dominant forensic process, as witnesses are asked to describe or identify suspects not known to them before the crime took place.

Much is known about the retrieval of faces seen once and for highly familiar faces, but the process by which an unfamiliar face becomes familiar is still a mystery. The difference between familiar and unfamiliar faces might be one of degree and not of distinct categories (Johnston & Edmonds, 2009). At one end of this scale, some faces are very well known to the participant, having encountered them over a long period of time on many occasions in different lighting and positioning conditions. The other end of this scale constitutes faces seen on one occasion for a very short time, and in one view. Consequently, variations in the way we process faces might be due to a difference in our experience of the face, that is, in our familiarity. Alternatively, there might be a qualitatively different way in which we recognize familiar faces to the way in which we recognize unfamiliar faces. Whatever the case, in their review of prior research, Johnston and Edmonds conclude that there are qualitatively different ways in which we encode familiar and unfamiliar faces.

One could argue that a person's face becomes more familiar the more we encounter it. However, Burton, Jenkins, Hancock, and White (2005) propose that the process of becoming familiar with a face is not a summation of our encounters with it, but rather an extraction of facial properties by eliminating those properties of the image that are not diagnostic of identity. If we encounter a face in only one instance, we are unable to know which of the characteristics we perceive will be crucial to representing the identity of the individual, forcing us to match the faces not by face properties but by image properties. Retrieval of unfamiliar faces might therefore be based on pictoral codes, literal mental representations of a particular picture, rather than abstract mental representations of the same face, so-called structural codes (Bruce, Burton, & Craw, 1992; Valentine, 2001). If novel viewpoints are presented at retrieval in a recognition experiment with unfamiliar faces, accuracy of recognition is reduced considerably (Bruce, 1982). Clutterbuck and Johnston (2005) and Young, Hay, McWeeny, Flude, and Ellis (1985) showed that photographs of different views of the same face are matched more quickly for familiar than for unfamiliar faces, suggesting a more view-independent presentation of familiar faces. An interesting approach was taken by Burton et al. (2005) to test different performances found for familiar and unfamiliar faces in a recognition task involving principal-component analysis. They compared the recognition of familiar face presentation, either of 20 images of one person morphed into one image, or a single image presentation. Participants recognized the depicted person better the more images were used to generate an average of this person's face. The authors argue that averaging multiple images of the same face produces a more reliable representation, thus reducing recognition errors.

The belief that a face, seen most likely only once and very briefly, is sufficiently detailed in memory to construct a good likeness of this face might be insufficiently challenged (Davies & Valentine, 2006). Research testifying a stable ability to recognize a face, even after years, might focus on processes related to the recognition, not recall, of a familiar, not unfamiliar, face. Familiar and unfamiliar faces might be encoded in different ways, resulting in a difference in recognition ability. For example, for familiar faces, the value of internal and external cues is reversed, with internal features being more significant for recognition (Bruce & Young, 1998). Memorising a face takes repeated exposure and prolonged encounters under different viewing conditions (Bruce, 2003). Unfamiliar faces are often recognized by their outer features, like hair style and face shape (Davies, Shepherd & Ellis, 1979). Their description, even while the face is in view, is selective and incomplete (Ellis, 1992). If an unfamiliar face is seen for a fleeting moment, and in a pose different to the one accessible in the composite system used, it might be that producing a composite imposes an unrealistic burden on the witness. However, it might well be possible. Surprisingly high recognition rates even after years or very brief encounters show that the facial information is stored in our memory, although it might be difficult to access.

2.2 Context-Reinstatement and Memory for Faces

Encoding Specificity

The theoretical principle of encoding specificity, which was first named by Tulving and Thomson (1973), states that an overlap of perceived stimuli between encoding and retrieval situations can aid in the retrieval of other, related information. This principle has been applied in many areas of research on memory, particularly to enhance the memory performance of eyewitnesses, where reinstatement of context is used as a retrieval cue to facilitate recall or recognition (e.g., Davies & Christie, 1982, 1988; Thomson, Robertson, & Vogt, 1982). According to Tulving (1985) and Tulving and Schacter (1990), encoding specificity can account for a diminished retrieval ability due to a mismatch between the way a stimulus is encoded and retrieved. Transfer-inappropriate processing (Meissner et al., 2008), on the other hand, suggests that separate memory systems are responsible for visual and verbal processing of faces. A change from holistic to featural, or from visual to verbal processing might result in the disruption of retrieval-based processing, diminishing performance.

Liu and Chaudhuri (2000) use the principle of encoding specificity to explain the failure to recognise an unfamiliar face: performance is predicted by the congruity between stimulus at encoding and stimulus at retrieval. Following this thought, the images of the person's face, shown at encoding and retrieval stage, might show too little visual compatibility to trigger recognition. A change in lighting, solution or image size between learning and test conditions can, as discussed before, cause an incongruence of the pictoral codes of the images, and therefore a failure to recognize. Supporting this hypothesis, Burton et al. (1999) found that face recognition was severely impaired when participants had to match poor-quality video clips with high-quality photographs when unfamiliar with the target people. Participants familiar with the faces of targets in the video clips performed significantly better over two experiments. Burton et al. concluded that face recognition for familiar people is less dependent on the current overlap between two perceptions of faces.

Aspects of Context Reinstatement

The term 'context' in studies relating to context reinstatement can be a 'conceptual garbage can' (Smith, Glenberg, & Bjork, 1978). Often within one study very different context components are confounded, making it impossible to distinguish between important and unimportant, that is, influential and non-influential and therefore superfluous context cues (Sporer, 2007). Contextual dependency of memory is influenced by a variety of encoding and retrieval conditions: the type of material learned, the type of context, the type of task during retrieval, and instructional variables. If, for example, an apple lies in the corner while a robber robs a store and its customers, showing that apple might not be an effective aid in retrieving more information. However, if that same robber throws over a bookcase, showing that bookcase during retrieval might have a higher chance of improving the memory for the event. Thus, if the robber interacts with aspects of the environment, these aspects will most likely have a greater influence when the environment is reinstated (Smith, 1979).

The physical surrounding of an event can be anything from the location, the size of the room, the objects and people present, odours, sounds, temperature or lighting (Smith, 1979). It is interesting to see that after 30 years, this original idea of global context has often been reduced to a purely visually and verbally perceptive context, with videos being used as stimulus material rather than a real event with life involvement and interaction between participant and situation or target (Dando, Wilcock, & Milne, 2009). In laboratory studies on face recognition, clothing as physical cues to the target face, shown before the identification task, increased recognition performance (e.g., Sporer, 1993a; Thomson et al., 1982). Likewise, field experiments on the effect of physical context reinstatement showed an increase in recognition performance when items with which the target interacted were re-presented (Krafka & Penrod, 1985; Smith & Vela, 2001), and when other aspects of the target's appearance were shown prior to the recognition task (Cutler, Penrod, & Martens, 1987).

For using context reinstatement in recall, a subject can make use of various sources as cues, the physical environmental being one of them. Those cues have a higher probability of being effective, since no other cues are imminent in the retrieval task (Smith, 1979). In contrast to that, in a recognition task, the participant is supplied with the highly effective, explicit cue of the target face, thus reducing her dependency on other contextual information. It might therefore be more difficult to achieve an effect of context reinstatement on recognition performance than it is on recall performance.

Cognitive Interview as one Application of Context Reinstatement

The cognitive interview is the most prominent application of the encoding specificity principle in mental context reinstatement for the purpose of enhancement of eyewitness' memory retrieval. The cognitive interview was originally developed as a guided retrieval technique to help the witness recall more information (Geiselman et al., 1984). Half of the interview mnemonics of the Cognitive Interview attempt to increase the likelihood of overlap between encoding and retrieval contexts, as described earlier in this thesis. This includes reinstating the environmental and psychological contexts of the event and reporting all information that comes to mind, including fragmentary information. Two more mnemonics encourage the witness to search through memory using a variety of retrieval paths, including recounting the event from a variety of orders and from a variety of perspectives. In the revised interview, two more aspects were included: mental imagery and rapport-building (Fisher & Geiselman, 1992; for reviews, see Koehnken, Milne, Memon, & Bull, 1999; Malpass, 1996; Memon & Higham, 1999). Fisher and Geiselman see two principles of the memory process as proven enough and crucial to support the application of their form of interviewing. First, a memory always consists of several elements, and some of these elements given as a retrieval cue add to the retrieval of the lacking elements. Second, one specific element of the memory in fact has many access routes. It might therefore be possible to reach this element through particular retrieval cues, even though other retrieval cues were not successful.

There are, however, important differences in the processes involved in recognition and recall. Thus a technique developed to enhance the number of details recalled might not be effective for enhancing recognition performance. Likewise, a description elicited with the aid of a Cognitive Interview might have an effect on later recognition of the target, or quality of a composite constructed. Several studies investigating the influence of the Cognitive Interview on recognition performance found no detrimental effect (Fisher & Quigley, 1990; Gwyer & Clifford, 1997). However, it is unknown to which degree the facial description of the target was emphasized in these studies, which could lead to a change in the cognitive processing of the target face. Finger and Pezdek (1999) found that when focussing the interview on facial information, the Cognitive Interview markedly impaired recognition performance. Schooler and Engstler-Schooler (1990) argued that when describing a face, participants are forced to engage in verbal-featural processing, thus interfering with visual-configural processing needed for a subsequent recognition task. However, follow-up experiments by Finger and Pezdek (1999) showed that by inserting a delay between facial description and subsequent recognition, this interference of the Cognitive Interview was neutralized, suggesting that the verbal description of the target does not overwrite the original presentation, but rather makes it temporarily less accessible. Meissner et al. (2008) also note that the vast majority of studies finding an effect of facial description on later recognition exclude subjective, more holistic dimensions of the face, but rather focus on featural aspects for which the participant is required to give a description. Studies by Brown and Lloyd-Jones (2005, 2006) support this notion by not finding a correlation between the number of holistic features described and the subsequent recognition accuracy.

Koehn et al. (1999) revised the original Cognitive Interview to be used with a visual instead of a verbal retrieval task, that is, the recognition of a target face. In order to avoid a potential effect of verbal processing on later recognition (Schooler & Engstler-Schooler, 1990), Koehn et al. promoted pictorial processing and tried to minimize verbal processing. Also, Koehn et al. encouraged witnesses to search for trait judgements they made while seeing the target face. Judgements and labels assigned to the target at encoding seem to be useful contextual cues at the retrieval stage for recognizing the target face (Chance & Goldstein, 1979).

According to Dando et al. (2009), police officers in England and Wales are currently asked to use the Cognitive Interview procedure when interviewing cooperative witnesses or victims. The ACPO Working Group for Facial Identification has, according to Gibson et al. (2009) highlighted the importance of the Cognitive Interview as developed by Fisher and Geiselman (1992) for a composite operator, with the Cognitive Interview constituting more than 50% of the training syllabus for those operators. A reason for this might be that E-Fit or Profit, the two programs currently used in the United Kingdom, require an initial description of the target face for the selection of the starting face, and a reduction of featural choices (Frowd et al., 2005a). Arguing against the use of the Cognitive Interview in composite construction, Gibson et al. consider the Cognitive Interview a process that tires out interviewer and interviewee, and deviates attention and concentration from the central task of developing a face composite. They suggest that only systems should be used that do not require an elaborate initial description.

2.3 Summary – Face Retrieval for Composite Construction

In using current composite systems within the police force, two assumptions are made about the witness's memory for the suspect's face. Firstly, a witness remembers all aspects of the suspect's face in a way that allows her to satisfactorily reconstruct the face. General studies on memory however show that memory is not comparable to a tape recording, and witnesses might not remember all aspects of a face, either because their exposure to the face was limited, because their attention to the face was diminished, or because they subsequently forgot the face. Secondly, the witness is able to relate individual features to her memory of the suspect's face. The theories discussed earlier emphasize the importance of configural information and the overall holistic impression, specifically in human faces.

Bruce (1988) emphasises the fragility of the memory for a face previously unfamiliar to the witness, seen in the offence under sub-ideal perceptual and emotional conditions, on which a composite can be based. In this situation context-reinstatement can be particularly useful, since according to Krafka and Penrod (1985) reinstatement of the original context is more effective in situations in which either suboptimal encoding or degraded encoding makes the accessing of memory difficult. Context-reinstatement can, as discussed earlier, help in recalling more details and, to a lesser extent, aid in the more accurate recognition of the target face.

Witnesses are always affected by the crime, irrespective of whether they had to suffer through it or witness it (Hinkle, 1990; for a review of the effect of stress on memory see Deffenbacher et al., 2004). To assume that the composite system is the only influential aspect that police can manipulate to achieve better quality composites might be imprudent. Yet, the only police force known to dedicate half their training time to the interviewing of witnesses before constructing a composite is the British Police Force (Gibson et al., 2009). Even this police force is, according to Frowd et al. (2008), being trained in the Cognitive Interview as introduced by Fisher and Geiselman (1992), rather than an adapted interview technique that considers aspects such as a verbal overshadowing effect or the different needs of context-reinstatement on recognition performance (Koehn et al., 1999).

Taylor (2001) claims that the way in which the witness is interviewed, irrespective of whether the system used is an artist, a mechanically assembled or a computer-based system, is of major importance. In a survey of US-police forces, McQuiston-Surrett, Topp, and Malpass (2006) found that no standardisation of any kind in interviewing the witnesses prior to constructing the composites was implemented, even though more than 60% of respondents elicited some form of facial description prior to composite construction. Taylor (2001) promotes the standardisation of interview procedures in order to obtain a face composite, hoping that this may lead to more professionalism and acceptability of the composite as evidence. Yet, it is not possible to standardise the procedure completely, since responding to the specific witness is a crucial part of a good interview. Nevertheless, loosely structured guidelines and awareness of possible successful components of interviewing might be possible. This might incorporate a rapport-building phase as described in the Cognitive Interview, suggested by Hinkle (1990), and teaching an awareness of the importance of not suggesting details or features, which might contaminate the original memory (Meissner et al., 2001) at the very least.

The question of which composite system is best remains unanswered, although certain assumptions regarding prior descriptions and incorporation of holistic or configural elements have been enunciated. All composite systems seem to require an initial facial description of some degree, be it basic descriptors for a choice of the initial subgroup (Gibson et al., 2009), or an elaborate description of facial features for a pre-chosen subset of available features (Frowd et al., 2008). The extent to which those descriptions disrupt the naturally occurring face retrieval process remains to be researched. Regarding the choice between different feature-based composite systems, prior research strongly suggests that witnesses need to be able to manipulate features within a face rather than make discriminations based on the isolated presentation of feature catalogues (Davies & Valentine, 2006). In addition, the possibility to store more facial features and to make more sophisticated alterations to the composite also suggest the departure from mechanical towards computer-based composite systems.

Chapter 3

Face composite research

Research into face composites most commonly focuses on various cognitive aspects and tools related to the construction process, investigating the premise that one experimental condition will achieve a better-quality composite than another. Past research on the enhancement of composite quality through specific systems, such as the use of context-reinstatement aids etc., has been summarized and interpreted outstandingly by Davies and Valentine (2006), and the interested reader shall be referred to this source for detailed information on quality-enhancing factors. To concentrate primarily on the aspects of the domain of face composites that are relevant for the studies presented in this thesis, this chapter will focus on how research can establish a quality-enhancement on composites under certain experimental conditions. In other words, I will discuss the different experimental designs employed within the domain of face composite research and their respective influence on the research presented in this thesis.

3.1 Composite Construction

Procedural Stages

Experiments regarding composite construction usually consist of two stages: A composite-construction stage and a composite-evaluation stage. Different, for example, from research on line-up identification, research on composite construction entails two sets of participants: participant-witnesses and participant-judges. Participant-witnesses play the role of witnesses of a crime, say for instance a robbery, and subsequently construct a composite from their recollection of the robber. The experimental manipulation constituting the independent variable is introduced during this first stage, either during exposure or composite construction. A second set of participant-judges assesses the quality of the composite constructed by various means. This subchapter will discuss the first stage of composite construction. The evaluation of composite quality will be discussed at a later stage.

In the composite-construction stage, a participant-witness is exposed to a target face, and subsequently has to construct a composite, a visual likeness of this face, created on the screen or scratchpad. As is the case in real life, the witness constructs the composite predominantly from her memory of the target face. Witnesses here have to rely on a facial processing that enables them to recreate a face they have most likely only seen on one occasion. As discussed in Chapter 2, it is cognitively demanding to extract facial properties from an unknown face, and to identify those characteristics that are diagnostic to identity. In other words, the participant-witness, while encoding the face in an experiment, has to filter out the characteristics of the face representing the identity of this specific person. Retrieval of this unfamiliar face might be based on the mental representation of an image rather than on the face itself, especially if a single photograph is used as stimulus material in an experiment (Bruce, 1982; Valentine, Darling, & Memon, 2007). A composite constructed with this mental representation in mind might therefore display characteristics of the image rather than the face. Consequently, the presentation of a stimulus in an experiment should ensure that the target face is encoded in a way that will enable the participant-witness to construct a composite of this target person from memory.

The encoding strength for participant-witnesses, in an experiment on composite construction, was tested by Koehn and Fisher (1997) among others. The researchers utilized a line-up twice: They assessed memory for the face from participant-witnesses, and composite quality from participant-judges. Participant-witnesses were asked to identify the target face after composite construction to estimate how well the face they had just constructed was encoded. Recognition accuracy was found to be almost perfect for participant-witnesses, but dropped considerably for the target-present condition in the composite evaluation. Comparing recognition rates for different groups might give us an idea of how much information we gain (or lose) by using composites as basis for identification, thus making a measure of recognition in participant-witnesses advantageous in two ways: Firstly, by providing an estimate of encoding strength and secondly, by generating a measure of task difficulty when using line-ups to test composite quality.

Composite Systems in Experiments

The independent variable being investigated most frequently in experiments is the ability of different systems to construct composites (Davies & Valentine, 2006). Since all composites are constructed with a system, be it mechanical, computer-based, eigenface-based or a scratchpad, the system is of considerable importance for composite quality. Relating the improvement in composite quality to the improvement of composite systems, Clifford and Davies (1989) refer only to the increase in representativeness in features available. However, in more recent years advances in systems also incorporated the way in which composites are constructed. As discussed in Chapter 2, new developments centred around the configural details of a face and on the holistic processing of faces. Davies and Valentine noted that software programs improved over the years by including a bigger variety of features, by the increased options to place and manipulate these features, and by adding graphic packages to the kits.

Testing this premise, Davies, van der Willik, and Morrison (2000) compared a mechanical system (Photofit) with a computer-driven system (E-Fit), finding that the newer E-Fit was superior only when composites were constructed in view of the target, and only on the measure of familiarity, a quality evaluation whose forensic relevance might not be immediately apparent. Likewise, Mag-a-Mug Pro, another computer-driven system, was found to be superior to other systems when composites were constructed in view of the target face (Cutler et al. 1988; Wogalter & Marwitz, 1991). Davies et al. (2000) suggest that the greatly expanded libraries of different features as well as an increase in flexibility of manipulations might be the underlying factor for this effect, but that this increased potential could not necessarily be harnessed when relying on the witness's memory.

Comparing pre-assembled systems to artists' drawings, Davies (1986) and Laughery and Fowler (1980) found an increased likeness rating for sketches, as well as an advantage for sketches when constructed in view of the target face as compared to composites constructed from memory. Comparing sketches to more modern systems, Frowd et al. (2005b) found artists' sketches being matched to target photographs as often as composites derived from computer-driven systems, but identified as a known person less often than through those systems. In a different study utilizing a longer delay, Frowd et al. (2005a) found an advantage of sketches over all other systems when a naming and matching task was used to determine composite quality. However, as mentioned in Chapter 2, training and professional practice of a sketch artist is in no way standardised, thus a conclusion that artist sketches are generally superior cannot be drawn. As stated by Davies and Valentine (2006) in their review of composite systems, any new software system or group of programs has yet to bring forward proof of its superiority– adduce evidence that it can outperform other systems in any of the quality evaluation measures, replicating this result over different experimental conditions.

Composite systems are commonly operated by a police officer or, in the case of experiments, a researcher, who might influence the resulting composite quality (Brace, Pike, Allen, & Kemp, 2006). Davies, Milne, and Shepherd (1983) and Gibling and Bennett (1994) found that the skill and experience of the composite operator is of significant influence to the ensuing composite. Subsequent research, taking the seeming superiority of experienced operators into account, utilized operators that were trained in the system used (e.g., Brace et al., 2006) or police officers with professional experience in composite construction in the field (e.g., Frowd, 2005a, 2005b). Designing a study on composite construction, any researcher has to take into account the influence of both the system and the person operating this system.

Experimental Design and the Independent Variable

The independent variable in experiments on composite construction is usually introduced during the stage of composite construction. Most variables being researched, like the system used or the interview conditions, are in one way or another present in every experiment. For example, studies use different composite systems to investigate ensuing composite quality. However, if composite system is not the variable of interest, a composite system has to be used that can produce good-quality composites, thus enabling the investigation of other variables, for example the effect of delay on composite quality. The influence of experimental design and independent variable on ensuing measurement of composite quality will be discussed in this chapter. For an extensive discussion of experiments investigating cognitive influences on composite quality, the interested reader shall once again be referred to Davies and Valentine (2006).

The delay between encountering the target face and constructing the composite from memory is one key aspect of experimental design. British police guidelines, according to Frowd et al. (2005a), aspire to a maximum delay between the incident and composite construction of 36 hours. In many studies, the composite was constructed immediately after the target face was shown to the participant-witness (e.g., Brace et al., 2006a; Bruce, Ness, Hancock, Newman, & Rarity, 2002; Davies et al., 2000). On the one hand, this practice seems driven by practical constraints – it is easier to administer the first stage of composite exposure and construction within one experimental session instead of having to ask the participant to come back for a second appointment. Moreover, Davies et al. (1978a), Green and Geiselman (1989) and McNeil et al. (1987) did not find greatly deteriorated composite quality after even a week's delay. However, these results might have arisen due to the rather low quality of composites in these studies in general. On a more general note, Meissner et al. (2008) see a clear need for a longer delay in eyewitness studies when wanting to draw conclusions about forensic settings since a longer delay might be the more realistic condition.

Investigating the influence of independent variables on the ensuing composite is not as straightforward as it might seem at first glance. Establishing cognitive influences of experimental conditions such as the use of different systems, the implementation of a certain delay or the addition of a context-reinstatement interview, is to a large extent related to the measurement of this influence. Experiments therefore have to rely on established composite quality measurements such as recognition rates or similarity ratings. In other words, a change in composite quality through the implementation of longer or shorter delay requires the reliable measurement of this quality. As will be discussed later in this chapter, the quality measurement itself might influence the inference a researcher tries to draw on the effect an independent variable has.

Illustrating the difficulties in estimating the influence of a cognitive manipulation on composite quality while constructing composites under forensically realistic conditions, Frowd et al. (2005a, 2005c, 2007a, 2008) conducted a series of investigations into the use of the Cognitive Interview. Frowd et al. (2005a) argue that a delay of 3-4 hours is desirable in the administration of the composite construction stage in an experiment, but not desirable when trying to achieve a forensically-sound realistic design. They argue that guidelines of the British police forces, as stated earlier in this chapter, should be implemented into the experimental design. Subsequent studies (Frowd et al., 2005b, 2005c, 2007a) contained a delay between encoding and composite construction of two days, as stipulated in the guidelines. In two studies, Frowd et al. (2005c, 2007a) found no overall effect of a holistic context reinstatement interview on sorting and naming accuracy. In a third study, Frowd et al. (2008) used a 3-4 hour delay, and found an effect of a holistic context reinstatement interview on the very same measures. Frowd et al. concluded that the holistic interview “offers the possibility of substantially improving the identification of criminal suspects” (2008, p. 276) and remarked that the technique is already being implemented as part of police interview training. Relating this to the demand for replicating an experimental finding under different conditions before promoting its implementation into policy and practice, one could argue that two studies with post-hoc comparisons or main effects bordering significance and only one study with a clear significance over all quality evaluations is not sufficient to implement this finding into general practice. It is clear that the holistic interview should not be overlooked but further investigation is required to determine the extent of its effectiveness.

Presentation Mode and Context Reinstatement

A variety of presentation modes has thus far been used to aid the encoding of the target face. As stated above, the mode of presenting the target face might influence encoding strength and subsequent composite quality, and should therefore not be overlooked. In experiments on composite construction, participant-witnesses are either exposed to a composite (e.g., Ellis, Shepherd, & Davies, 1975), a photograph (e.g., Christie, Davies, Shepherd, & Ellis, 1981; Davies et al., 1978a; Davies, Ellis & Shepherd, 1978), or witness a staged event, either live (e.g., Davies & Milne, 1985; Laughery & Fowler, 1980) or by viewing a video (e.g., Bruce et al., 2002).

An incentive of using moving images (through videos or live events) of the target face is to avoid biasing the outcome by using the same photograph during encoding and evaluation stage, as was demonstrated in the study by Davies et al. (1983). In this instance, the experimenter runs the risk of evaluating the picture recognition or likeness rather than the target face recognition or likeness, an important distinction in the evaluation of composite quality. As discussed earlier, encoding a novel face seems to be more prone to pictoral encoding rather than encoding of facial characteristics, and as such a single photograph runs the risk of encouraging this former type of processing. However, a picture-bias can also be avoided by for instance using several photographs during encoding and/or evaluation, as was done by Davies et al. (2000). Here, the target face was shown in three different poses during encoding, thus making the construction of a composite less dependent on qualities inherent to any one of the photographs shown at both stages. Alternatively, Gibling and Bennett (1994) used separate photographs of the same target face during encoding and evaluation stage, thus avoiding the measurement of picture similarity in contrast to facial similarity. However, as stated in Chapter 2, memory for an unfamiliar face is more dependent on the viewing conditions in which it was encountered. Hence, using two different photographs during encoding and retrieval might encounter the risk of making the composite unrecognizable.

A live event as implemented by Davies and Milne (1985), Koehn and Fisher (1997) and McNeil et al. (1987) bears another cognitive advantage beside a rich encoding context – that of unintentional learning. In the experiment by McNeil et al., participant-witnesses met and interacted with the target person on a false pretence, and were only informed about the nature of the task when having to construct a composite of that same person. Likewise, Koehn and Fisher had target persons instruct participant-witnesses about a bogus experiment, using a script of specified verbal and non-verbal interactions between each participant-witness and target. In doing so, the researchers attempted to have participants encode the situation and the target as naturally as possible, presuming that in a real-life situation the witness would not try to remember the target face for a later composite construction, but rather be asked to do so after the situation was resolved and the police had arrived. Arguing along the same line, Wells and Hryciw (1984), as stated in Chapter 2, concluded that naturally occurring encoding is more holistic, whereas telling participants that they will have to recall the face for composite construction might lead to a more feature-based encoding strategy.

Investigating the influence of context-reinstatement techniques, the use of live events or videos of events might have an additional advantage. As discussed in Chapter 2, it is presumed that the richer the context during encoding, the more interwoven the facial memory is to a set of associations, and the easier it might be retrieved for the purpose of composite construction. Malpass (1996) surmised that the bigger the overlap between retrieval environment and original environment, the better the recollection of the event will be, making an enriched encoding even more desirable for a difficult retrieval task like constructing a composite.

Many of the key studies utilizing context reinstatement techniques to enhance facial recollection used a live event as stimulus material, thus adhering to Malpass's premise of a rich encoding context. Watching a videotaped event can, however, also provide a sufficiently enriched context for later reinstatement (Cutler & Penrod, 1988). Using both a staged or a videotaped event, Koehnken, Milne, Memon, and Bull (1994) found that the staged event resulted in the retrieval of more recalled details after administering a cognitive interview than the videotaped event. Consequently, when researching context-reinstatement interview techniques, measuring the influence of an interview on composite quality might benefit from a rich encoding context, as can be found in a video and, more so, in live events.

3.2 Measurement of Composite Quality

What is a Good Face Composite?

The judgement of what a good-quality composite entails is necessary for measuring the outcome of experiments as well as police investigations. The composite itself is the outcome of the construction process, that is, the product of the memory of the witness, filtered through the system operator and composite system. However, judging what a good-quality or a bad-quality composite is appears to be a complex and challenging task. Raising awareness of all aspects of this problem is the aim of this chapter and at least partially solving it is the aim of this thesis.

Early studies measuring composite quality focussed for example on aspects like physical resemblance. Ellis et al. (1975) estimated composite quality by counting the number of composite features being used both in the stimulus composite and the constructed composite. However, this approach focuses solely on featural quantities, disregarding similarities between individual features as well as configural and holistic dimensions as discussed in Chapter 2. Taking both featural and configural information into account, Rhodes (1988) measured composite quality by measuring the Euclidean distances such as length of nose, width of mouth. Again, holistic dimensions, proposed to have a dominant influence in face processing by Hancock (2000), are disregarded in this approach. Several currently used evaluation methods focus on the recognition of the face as a specific person, as will be discussed in more detail later in this chapter. This approach defines quality as recognizability. Many more approaches exist towards measuring composite quality, and there is no consensus on how quality can be defined, let alone measured. In other words, there is no cognitive or forensic definition of which quality a composite should possess to be regarded as good or bad.

The field of face composite construction is not the only research area struggling to define the accuracy and recognizability of a facial image. In the area of facial approximation, that is, the method of building a facial representation from a human skull, accuracy assessment methods are considered to be responsible for disparate results found in studies using different methods (Stephan & Cicolini, 2008). The authors used a likeness rating task and a recognition task to assess the same facial approximation, and found that a lack of calibration made a comparison across methods difficult if not impossible. The authors also used foil faces for additional likeness ratings towards the facial approximation, and found that scores were similar irrespective of the compared face. In other words, likeness ratings for the foil faces were as high as those for the facial approximation. The authors concluded that likeness ratings are an insensitive measurement for accuracy, and that familiar and unfamiliar face recognition tests should be given precedence over rating tests.

Absolute or Relative Quality of Composites

The likeness of a composite to the target face is sometimes described as having 'little resemblance' or 'reasonable resemblance' (Brace et al., 2006a). Those qualifying adjectives suggest an absolute scale with which composite quality can be measured, yet in my understanding no scale or baseline allowing for an absolute judgement has been established. In order to have an absolute instead of a relative measure of composite quality, one has to define the lower as well as the upper boundaries of this scale.

Measuring memory retrieval has a long-standing history. Early approaches like that of Ebbinghaus (1895, as cited in Koriat, Goldsmith, & Pansky, 2000, p. 482) calculated the number of correctly retrieved input items, thus being able to give an absolute estimate of retrieval accuracy. The number of correctly retrieved items relative to the total number of items constituted the percentage of accuracy. However, estimating the quality or accuracy of a composite is more difficult. As described earlier, there is no specific number of equally weighted features that, being correctly assembled, constitutes a whole face. Even taking only a featural viewpoint, two different noses, for example, with one being correctly and one incorrectly assembled, can still be sufficiently similar to make the face recognizable.

Defining lower and upper boundaries of a potential scale might be equally challenging. Composites were originally intended to replace facial description, as the latter is considered to be problematic. Witnesses might lack the linguistic skills to describe a face appropriately (Hinkle, 1990; Penry, 1971; Taylor, 2001), and might lose the memory for the original image by transferring a visual image into a verbal description (Meissner, Brigham, & Kelley, 2001). Following this thought, a lower boundary against which composite quality could be measured, is the increase in recognition or likeness rating of the composite in relation to a given description (Koehn et al., 1999). However, studies often relate quality measures to chance level. 'Chance level' here refers to participant-judges identifying the target from the composite no better than if they had made their choice randomly. Stephan and Cicolini (2008) introduce a different concept of expectancy level, comparing composite-target choices against composite-foil choices. In summary, any of these approaches, the comparison to a verbal description, the comparison to foil faces or the use of chance level, might be appropriate in setting a lower limit or baseline.

Regarding what quality to expect, Bruce et al. (2002) state that recalling faces is an inherently difficult process, thus the construction of a composite is likely to hold some distortions relative to the true appearance. Translating this into the measurement of for example recognition accuracy, this statement implies that a perfect or close-to-perfect recognition rate is an unrealistic expectation. Defining a possible upper limit for composite quality, Koehn and Fisher (1997) introduced an 'ideal' composite into the study design. This composite, constructed by the operator while viewing the target face, serves as a standard against which composites constructed under other experimental conditions are evaluated. However, in most studies, there is no rationale to define an upper limit for a composite quality scale. Subsequently, most studies on composite construction entail the judgement of relative quality rather than absolute quality, stating that a composite is better or worse than another, rather than stating that a composite is good or bad.

Single or Multiple Dimensions of Composite Quality

It is common to use several measurement tasks to estimate composite quality. This approach seems both necessary and difficult. It seems necessary because composite quality seems to consist not of one but of many dimensions. Studies measuring composite quality measure accuracy and usefulness, subjective and objective quality, recognizability and similarity. For example, a composite could be of good likeness but still not be useful to the investigation if no one who sees it on television or in the newspaper actually knows the suspect. Likewise, a composite that is of fairly poor likeness might still bear enough of a resemblance to the suspect to result in calls from members of the public familiar with them. In experimental terms, composites constructed with a specific software system might show an increased recognition rate, but no such increase when employing a similarity measure. The difficulty in employing multiple measurement tasks for composite quality lies in the fact that multiple measurements often have different outcomes. Each outcome subsequently requires an interpretation regarding which qualities of the composite have been measured, and which qualities are important.

Up to three evaluation measurements are used in the composite evaluation stage of each individual experiment. Multiple measurement tasks often lead to different measurements having different outcomes. Frowd et al. (2005c) for example, used three measurements of composite quality: Naming of famous celebrities, matching composites to target photographs, and cued naming, that is, naming of composites followed by the viewing of a list of those targets from which the composite was constructed. Frowd et al. found floor-level results for the naming task, a task described as being the most forensically relevant quality measurement. In the matching task, described as feature-to-feature in-view comparison, the researchers found an average sorting accuracy of 50%, but no overall significant difference between experimental conditions. Lastly, when giving participants a list of celebrity names from which to choose the target who might be depicted in the composite, a significant difference emerged between interview conditions. One could argue that undue task difficulty in the naming task was responsible for the floor-level performance, making it impossible to find differences between experimental conditions even if they existed. On the other hand, one could argue that choosing one target from a sample of eight given target photographs or names is not a forensically valid measurement. Frowd et al. dismiss the outcome of the matching task, stating that a from-memory comparison between the composite and the target face, for example in a naming task, is the more viable measurement to establish recognizability of the composite. In the end, placing value on the different outcomes currently seems to be more a matter of individual interpretation rather than being founded in an agreement among researchers.

Relating the different quality measurements to the intended forensic utility of composites, Koehn et al. (1999) suggest that the value of the face composite is the degree to which it enhances the investigation beyond that of using only a verbal description of the suspect, not whether the facial composite, by itself, can lead to a suspect or target identification. In other words, demonstrating that a face composite can lead to a correct identification is more of theoretical than of practical value. Contradicting this statement, Clark defined best practice of composite construction as 'the method that produced the most correctly identified composites' (2000, p. 9). These two statements highlight the need for a more thorough and clear definition of a 'good-quality composite'. The question that this thesis sets out to answer, at least in parts, is if a composite possesses distinct cognitive qualities, and if these qualities, be it a single quality or multiple qualities, can be measured.

Task Difficulty and Its Influence on Quality Measurement

Besides the need for a better definition of the concept of composite quality as discussed earlier, measuring composite quality requires a measurement of task difficulty. The assessment of task difficulty was discussed earlier, concluding that for analysing the absolute composite quality it is necessary to have an idea of which upper limit and lower limit can be expected. In other words, in order to state that a composite is of good quality because it is recognizable as a specific person, one would have to be able to estimate how many realistic alternatives to this person exist, and what chance level recognition would be.

Pre-evaluating task difficulty could have the additional benefit of preventing floor-level results prior to the administration of an evaluation task, as was reported in studies by Koehn and Fisher (1997) for both their composite quality evaluation methods, and by Frowd et al. (2005c) for a naming evaluation, in which none of the composites were identified. Thus, while establishing task difficulty is desirable but not necessary to analyse relative instead of absolute differences between experimental conditions and the subsequent improvement of composite quality, it would have the advantage of avoiding non-interpretable results.

An anecdotal example is described by Frowd et al. (2005c) whereby participants were given a list of celebrity names which they had to match to composites. Included in the set of target faces was only one bald man, and composites depicting him achieved close-to-perfect identification. Frowd et al. concluded that this identification rate was only achieved by the hair style or rather the lack thereof, thus no conclusions could be drawn on composite quality for this specific target. Relating this to task difficulty, it would be fair to say that with only one option to choose from, high recognition rates were a direct result of low task difficulty.

Current Measures

Although a multitude of different composite quality evaluation techniques are used in the domain of face composites, five categories of measurements currently used can be defined, namely likeness rating, matching, naming, identification and mugshot technique. Each technique will be discussed below, relating it to cognitive concepts involved as well as differences in experimental designs.

Firstly, a likeness rating of the composite in the presence of the target face is considered by some researchers a feature-match evaluation. In other words, participant-judges compare composite and photograph, being presented simultaneously, on a feature-to-feature basis, and judging similarity accordingly. Rating the likeness between a composite and the target face is often described as a method to evaluate the subjective similarity (e.g., Koehn & Fisher, 1997) in contrast to objective likeness as measured by the Euclidean distance (O'Toole, in press; Rhodes, 1988). Ellis and Shepherd (1992) describe a case in which a composite was drawn by a victim of sexual assault, and a custody officer saw likeness between this composite and a suspect being questioned for a similar offence, who was later convicted of both offences. Here, the composite's likeness to the target face, together with additional information, helped the police identify a suspect.

The most common likeness rating is one in which participant-judges rate all composites on their one-dimensional resemblance to the target face. This likeness rating is usually done by displaying, in a booklet or on the screen, a composite together with a full-frontal photograph of the target person, asking for a rating on a scale of for example 1-10 (Bruce et al., 2002; Koehn & Fisher, 1997). Deviations from this practice incorporate the display of several photographs in different poses (Bruce et al., Experiment 2) or rating the likeness with only the composite in view (Brace et al., 2006b). The outcome of a likeness task is one or several ratings for each composite, subsequently treated as interval data for further analysis.

Recognition of a familiar person from a composite is considered by some researchers as the most forensically valid measurement (Bruce et al., 2002; Frowd et al., 2005a, 2005b, 2008; Koehn et al., 1999), and is subsequently called naming task. A composite may be similar to, and therefore remind the person of someone he/she knows from previous encounters. An alternative use would be not the inclusion but the exclusion of certain people: A composite may appear so different from a known suspect that it excludes the suspect from further investigation (Koehn & Fisher, 1997).

Bruce et al. (2002) and Davies et al. (2000) used departmental staff members as target persons from which composites were constructed, subsequently asking students to identify composites as one of their lecturers. Alternatively, Brace et al. (2000, 2006a) an Frowd et al. (2007a) used celebrities as target persons. In most experiments, knowledge of the target's identity was evaluated by asking participant-judges to identify the target person in the photograph, after attempting to identify the composites. In all experiments using famous or known target persons, participant-witnesses were unfamiliar with the target person. This was achieved by either administering composite-construction in another country (Frowd, McQuiston-Surrett, Anandaciva, Ireland, & Hancock, 2007b) when using local celebrities, or by choosing participants from other universities when using departmental staff members.

Naming is often used by Frowd et al., and is said to achieve an average of 20% identification rate using a short retention interval (2004, 2005b, 2007a; also Davies et al., 2000), but drops to only a few percent when the retention interval is increased to more than a day (Frowd et al., 2005a, 2007a). Fluctuations in the naming rate might be due to differences between experiments in average composite quality, or differences in task difficulty. Task difficulty here might relate to the number of possible alternatives to the target person, that is, the population size the targets are drawn from. Frowd et al. (2005c) for example found no identification of any composite when instructing participant-judges to name 'well-known British celebrities', a population that might have a considerable size. Bruce et al. (2002) used a more specified target group for their evaluation task, asking participants to identify a member of the department the participants were studying or working in. In this case, it would have been possible to specify the number of possible alternatives to the target face, thus determining an estimate of chance level rate for correct naming.

Supporting the notion that naming rates might be dependent on the population from which the target persons were drawn is a study by Kovera et al. (1997). They found that former high school students could identify only 6% of composites depicting former classmates. The authors concluded that their findings "raise doubts about the likelihood that composites prepared under field conditions will yield a pinpoint identification of a perpetrator by individuals who know the perpetrator" (Kovera et al., 1997, p. 245). However, as discussed in Chapter 2, a pinpoint likeness might actually not be an achievable goal.

Alternative aspects of experimental design were implemented by Brace et al. (2006a), who administered the naming task in groups, asking participants to identify the composite of a famous person they should be familiar with. The authors differentiated between incorrectly named composites and composites not named, thereby enabling an analysis of incorrectly named composites. Incorrectly named composites might be one way to estimate the number of possible alternatives to the target person within a population.

The recognition of an unfamiliar face as the person depicted in the composite is implemented into experimental design by a line-up, a mugshot or a matching task. In all three tasks, participant-judges have to decide if the person depicted in the composite is presented in the photo-array of faces displayed, and if they can guess which one it is. Since recognition of a familiar face might involve different processes than recognition of an unfamiliar face, it might be worthwhile to distinguish between these two evaluation methods (Johnston & Edmonds, 2009). In a real-life setting, police officers might use the composite to compare it to possible suspects in a case. For example they can decide between felons convicted of similar offences, which one of them might warrant a closer investigation. Alternatively, a composite placed on a notice board might prompt a shop employee to report that she has seen a similarly looking man during her shift, on the day in question, in the same area.

The sorting or matching task is a forced-choice identification of an unfamiliar face. Participants are commonly presented with an array of different photographed target faces, simultaneously laid out in front of them. For each composite, participants then have to match the person depicted in the composite to the most likely candidate, thereby forming piles of matched composite cards next to each photograph (e.g., Davies & Milne, 1985). The outcome measure here is hit or miss, that is, the correct or incorrect matching of composite to target.

Line-up measures are sometimes considered a direct measure of composite usefulness and utility (Koehn & Fisher, 1997), and as more forensically valid than a rating or matching task (Green & Geiselman, 1989).

Correct identification rates as estimate for recognition accuracy across experiments utilizing a line-up can vary widely; ranging from as low as 8% (Parker & Ryan, 1993) to as high 80% (Pozzulo & Lindsay, 1999). Clark, Howell, and Davey (2008) found an identification rate of 46% for target-present line-ups, averaged over 94 studies, when studying recognition memory. Variability in results can be expected due to variations in experimental design. In spite of this, a meaningful estimate of absolute recognition accuracy that relates to the recognizability of the face or composite across experiments would be desirable. To discriminate more clearly among the range of cognitive influences on recognition accuracy, the framework of Signal Detection Theory (Green & Swets, 1966) can be used to separate the distinct contribution of memory and that of decision criteria on the identification rate in a line-up. In having both target-present and target-absent line-ups, a researcher can estimate if the memory is more accurate under a certain experimental condition, or if this specific condition makes the participant choose more readily without increasing accuracy (e.g., Clark, 2005; Malpass & Devine, 1981). The concept of recognition accuracy as being different to the willingness to choose is called diagnosticity (Wells & Lindsay, 1980; Wells & Olson, 2002).

Consequently, composite diagnosticity could be considered an intermediate measure of the memory strength of participant-witnesses, and the translation of this memory into the composite's appearance. Relating both concepts to composite quality, a composite can be identified as a specific target in a line-up because it shows distinct features, configurations, or a general impression that is related only to this target person, resulting in a higher composite diagnosticity. On the other hand, a composite can yield a higher identification rate because it has a less individual and more average appearance, resulting in an increased willingness to choose a person from the line-up. Both concepts can only be measured if both target-present and target-absent line-ups are used, and when the participant is permitted to reject the line-up.

Utilising this approach, Koehn and Fisher (1997) for example found an increased rejection of the line-up in both target-present and target-absent line-ups when administering a context-reinstatement interview, suggesting a decreased willingness to choose in this experimental condition. By contrast, Bruce et al. (2002) administered a forced-choice target-present line-up in their first experiment, thus making it impossible to estimate the different impacts of diagnosticity and willingness to choose on recognition rates. In their second experiment, Bruce et al. added target-absent line-ups to examine the relative effect of the different conditions on triggering hits as well as false alarms.

Again borrowing from the area of eyewitness identification research, line-ups can be pre-tested for effective line-up size as well as bias towards the target photograph (Tredoux, 1998). Prior examination of potential line-up bias was implemented for example by Gibling and Bennett (1994). The researchers showed pilot study participants a photospread with the target position being randomised, and asked if any of the photographs stood out. Alternatively, Bruce et al. (2002) measured composite quality using line-ups without prior measurement of task difficulty and bias against the target. However, Bruce et al. subsequently discussed only significant differences in recognition rates rather than relying on absolute rates. It is interesting to note that the authors reported a significant difference overall (target-absent and target-present condition combined). According to the rejection frequencies reported, participants significantly more often chose to reject the line-up correctly in the target-absent line-up and incorrectly in the target-present line-up ($\chi^2_{(1)} = 7.65$, p=.006). The authors noted that the higher rejection rate in the target-absent line-up shows the more protective nature of morphed composites against misidentification, thus an increase in discriminability. However, the increased rejection rate in the target-present line-up might encourage another interpretation: The morphed composites encouraged a higher response criterion rather than an increase in discriminability.

A different measure of identification is the use of mugshots. Christie et al. (1981), Davies, Shepherd, Shepherd, Flin, and Ellis (1986), and McQuiston-Surrett and Topp (2008) implemented this recognition measure. Participants are typically confronted with a simultaneous presentation of photographs of all targets and a varying number of foils. McQuiston-Surrett and Topp added one foil photograph for every target photograph, whereas both Davies et al. and Christie et al. added three foil photographs for every target photograph to the mugshot. The latter two experiments allowed participants to either choose one photograph or reject all photographs. However, considering the increased number of possible choices in this task, a second and third choice count as reported by Ellis, Davies, and Shepherd (1976) might more appropriately reflect the assumption that a composite should reflect a type likeness rather than lead to a correct identification of the target person out of a group or persons or photographs (Koehn et al., 1999).

3.3 Summary – Recommendations for Further Research

It is apparent from the literature discussed in this chapter that the question of the best approach to studying composite construction is highly contentious. It has been shown that many different approaches towards design aspects and appropriateness of design with regards to forensic implications exist, most of which have merits. However, all or only a few aspects might have forensic value.

It would require the close study of forensic reality to determine if policy or anecdotal remarks bear weight in describing police practice. In other words, we need to know more about forensic practice regarding composites to decide which experimental design, both in terms of composite construction and composite quality evaluation, is most appropriate to reflect police reality. I do not consider it sufficient to judge the forensic value of composite experiments on the basis of personal experience or official policy alone.

Taking the example of the delay period between viewing the incident and constructing the composite, several studies have shown that naming rates drop considerably from employing a very short delay to the Home Office recommendation of a delay not exceeding two days between crime encounter and composite construction. More specifically, recognizing a member of the public from a composite constructed two days after the incident was successful only in one out of thirty cases (Frowd et al., 2005a). It might be interesting to investigate if the delay period in real-life police work is even longer, and if so, whether the naming rate decreases even further after this extended period, or if it plateaus as suggested in Chapter 2. For now, it cannot be said how composites would fare after a delay of for example 1-2 weeks.

As a second aspect, the mode by which composite quality is evaluated warrants further discussion. As stated before, it is unrealistic to expect that a composite alone enables a person or participant-judge to recognize one specific person out of a population of hundreds of celebrities. Then again, if investigating officers publish a composite and receive twenty replies, one of which leads to the apprehension of the perpetrator, composites have to be recognized as having some practical utility. In order to determine if this is the case, more information is needed on the use and utility of composites in real life police investigations.

This chapter has illustrated the multitude of experimental approaches towards improving or measuring composite quality. One can distinguish between them based on the different cognitive concepts they rely on. For example, the comparison of composite to target person can be done in view of the target face or relying on memory of either one or both faces. Again, forensic utility of either of those comparisons has to be established before disregarding any one of them. Participant responses can be more relative judgements, as is the case in a forced-choice matching task, or more absolute, as Lindsay and Wells

(1985) claim is the case in a sequential line-up task. To further this distinction of cognitive aspects, we need experiments yielding similar results for tasks that rely on the same cognitive aspects. In other words, to support these cognitive interpretations of measurement tasks, an experimental design would have to be implemented in which different measurements are employed investigating the same composites with comparable means. A methodological comparison of composite quality measurements is however difficult, since past studies focussed on investigating independent variables instead of the measurement itself. A study designed to incorporate many commonly used measurement tasks would enable a beneficial comparison between those tasks and would thereby facilitate an important shift in focus.

Chapter 4

The Use of Context Reinstatement in Compositing

4.1 Introduction

In this chapter I start my empirical exploration of composite quality enhancement by focussing on the interview method used prior to or during composite construction. I hypothesize that if an interview method reinstates the original context in which the face was encoded, composite quality will improve as a result.

Using context reinstatement to enhance composite quality is not a new concept, although few studies have systematically focussed on the applicability of context-reinstating techniques in this research domain. As discussed in Chapter 2, these studies showed competing results. Derived from the concepts inherent in the Cognitive Interview (Fisher & Geiselman, 1992), Koehn and Fisher (1997) developed an interview form for face composite construction in which they used visual imagery aspects to help the witness keep the target face in mind while constructing the composite. They found no difference in composite quality when comparing their adapted cognitive interview technique to a standard police interview comprising questions about specific features. However, this result could also come about due to floor-level recognition and likeness ratings. Conversely, Davies and Milne (1985) found an effect of both mental reinstatement and physical reinstatement. Using a different software system as well as different composite quality evaluation methods, Davies and Milne found identification rates well above an assumed chance level. Since the software system and its ability to produce good-quality composites seem to influence the distinguishability between experimental conditions, two systems are used in studies in this chapter which showed the capability to produce good-quality composites.

A methodological issue concerning the application of a context-reinstatement interview prior to composite construction concerns the Verbal Overshadowing Effect (VOE, Schooler & Engstler-Schooler, 1990). Schooler and Engstler-Schooler found that verbalisation of facial features can impair later recognition. In other words, verbal processing of a visual image can impair the visual memory itself. Thus, visual processing should be promoted instead of verbal processing of faces when a witness or participant-witness has to identify a face. Consequently, Koehn and Fisher (1997) instructed witnesses to not express their memory of the target face, but rather to form a mental picture of the features in question before constructing a composite. On the other hand, Meissner and Brigham (2001a) determine that only in a limited number of situations the description of a face leads to its decreased recognition. Since face composite construction is in parts a recognition task, it is nevertheless prudent to assume that a verbal description of facial features could change the way in which the witness remembers the face. To prevent this alteration of memory due to verbal processing, participant-witnesses in my studies were asked to give only a short description of the target face as a starting point for the later construction. After this initial description, participants were given sufficient time for a possible VOE to dissolve (Meissner & Brigham, 2001a), and were asked not to verbalize any characteristic of the target face in the context-reinstatement interview.

Another methodological issue concerns the manner in which a context to a target face is constructed. Luu and Geiselman (1993) used a video of a staged crime, and found that the cognitive interview yielded an improvement in composite-quality when a more holistic composite system was used. Likewise, Frowd et al. (2008) found an effect of an adapted context-reinstatement interview on Pro-Fit composites when using a soap opera video clip. As discussed in Chapter 3, videos as well as live events seem to generate an encoding context that can be drawn upon for a subsequent context reinstatement interview. For this reason video clips showing a staged robbery were used in the first experiment.

Both featural and holistic software systems are currently used by the police for composite construction. It seems as of yet inconclusive whether holistic systems using whole faces (e.g., ID, EvoFit), or featural systems selecting each feature within the context of a face (e.g., Identikit, Pro-Fit) are a better tool for facilitating the witness's memory retrieval. Whichever system is used, it will influence the retrieval process of the face. An interview method which encourages retrieval of holistic face characteristics such as personality traits, as has been used in studies by Frowd et al. (2005c, 2008), might have a detrimental effect on retrieval when used in conjunction with a featural system (and vice versa). More specifically, encouraging holistic retrieval might be beneficial when constructing a composite with ID, making it unnecessary for the witness to remember a specific nose or eyebrow, as would be the case in a featural system. Taking into account that we do not know if a more holistic or more featural face retrieval is beneficial to yield better-quality composites, and that we do not know if a more holistic or featural outside influence on this retrieval will aid the witness to retrieve more correct information, I used both featural and holistic interview methods as well as software systems.

4.2 Study 1

Summary of Experiment

A 2 (Composite system: Identikit, ID) x 3 (Context reinstatement strategy: holistic, featural, none) between-subject factor design was employed. In the composite construction phase a video of a staged robbery was shown to participant-witnesses, from which they had to reconstruct the target's face after a two-day delay. Composite quality was subsequently evaluated by two measures: a rating of similarity between composite and target photograph, and an identification of the target from an array of six photographs. Composites and target photographs can be found in Appendix E-01[2], photo-arrays can be found in Appendix E-02.

[2]Please note that appendices are available in printed form to the back of the printed thesis (when labelled with a P), or in electronic form appended as a CD to the back of the thesis (when labelled with an E).

Composite Construction Phase

Participants

Eighteen persons were recruited by advertisement as participant-witnesses on the campus of the University of Cape Town. In order not to influence the results through the cross-race effect (Meissner & Brigham, 2001b) only participants who described themselves as 'White' were recruited. Each of them was offered ZAR 20 ((2.60 US$) for their participation. Since no staged live event was used, participants were informed about the purpose of the study, and gave their informed consent.

Materials

Three drama students at the University of Cape Town were chosen from a group of students who volunteered, for payment, to serve as mock perpetrators in a staged robbery. The targets were young white males (average age 20 years) without beard, accessories or unique facial marks (e.g., scars). This target group was chosen to account for possible age-related expertise of the participant-witnesses for faces of their own age (Wright & Stroud, 2002). Only White targets were chosen to be able to obtain a good quality likeness. There is a very limited range of Black and Coloured South African faces that is a part of the Identikit data set, thus hindering the participant-witnesses choice and construction of the composite, even though they might have a good memory of the target face. Also, students serving as targets were not attending courses on the main campus, and were unlikely to be known to the participant-witnesses. Three video clips were produced as stimulus material, showing the same staged robbery of a campus bookshop, with the only difference being the robber himself (see Appendix E-03, E-04, and E-05). The video showed the target entering the shop, taking a book to the counter, discussing the price with the female shop assistant, and running out of the shop with the book in his hands. A staged theft was used to create a more active and realistic context for later context-reinstatement as stipulated by Meissner et al. (2001). The video-clips were later shown on a computer to the participant-witnesses.

As composite system Identikit 2000 and ID (Tredoux et al., 1999; Tredoux, Nunez, Oxtoby, & Costa, 2007) were used in the experiment, both of which are described in detail elsewhere (Davies & Valentine, 2006). It is important to note that Identikit as a feature-based system contains a multitude of facial features (e.g., ears, eye brows, and chin) which are chosen from an array of similar features in isolation, and are subsequently altered in the context of the face under construction. The Identikit system is very popular, being one of the leading composite systems worldwide (McQuiston-Surrett et al., 2006) and the only system used by police in South Africa (Schmidt & Tredoux, 2006). ID is based on holistic face processing, and presents an array of computer-generated faces from which the participant has to choose the closest in likeness to the target face. With an eigenvector-based algorithm the software then calculates another set of faces closer to the chosen face. This procedure is repeated until a sufficient likeness is reached. For both systems the witness was assisted by a composite system operator (author), trained and experienced in each of the two systems equally well by self-study, on-the-job training by an experienced police officer as well as several trial runs prior to the experiment.

As experimental manipulation, an interview guideline was used mainly relying on context reinstatement (Krafka & Penrod, 1985; Memon & Higham, 1999). The interview was implemented before composite construction started since it had to be the same for different systems used. All participants underwent a pre-construction interview as a common technique for rapport building and obtaining necessary information (Frowd et al., 2005a; Taylor, 2001). This common guideline entailed a short rapport-building stage for the participant-witness to become adjusted to the situation, an explanation of the system and what they had to do. As a second stage all participants were asked to report race, gender and age of the target[3], without being encouraged to further elaborate their explanation by the system operator who conducted the interview. Depending on the condition, participants were either asked to start constructing a composite, or were given context-reinstatement instructions.

In the context-reinstatement conditions the operator explained to the participant-witness that first a context-reinstatement interview would be employed in order to help them remember the face of the target, after which they would construct the composite of the target person. The participant-witness in the context reinstatement condition was asked to close their eyes and think back to the video clip and try to form a mental image of the target. This procedure was

[3]These questions where a necessity to establish the database (ID) and the first suggestion for a composite (Identikit).

used in order to influence the participant towards a certain strategy of retrieval, not assuming that the participant in fact employed a context-reinstatement strategy. The same is true for the differentiation between feature-based and holistic strategy; only the effect of the different forms of interviewing, aimed at different internal retrieval strategies of the participant, could be manipulated. Subsequently, the strategy-related details given by the witness were repeated in the cued recall phase, and several categories of information were addressed. These were hair, eyes, nose, mouth, chin, hats, glasses, beard, complexion, height, built, weight, tattoos/scars and body defects for the feature-based interview, and language, origin, habits, guesses about personality (honesty, friendliness, excitability, selfishness and arrogance), intelligence and profession for the holistic interview. In both strategies the whole of the target person and not only their face, was used as the source of cues, not only the face of the target. All details were closely related to the central stimulus (target face). After the cued recall phase the participant was asked to visualize the face again in her mind, without describing or reacting to this.

Procedure

The participants were tested individually throughout and were randomly allocated to cells of the design. Participants made two visits to the laboratory. Upon their first arrival participants were told that they would watch a video of a staged crime, that they should picture themselves as eyewitnesses to this crime, and lastly that they would later be asked to report the events that transpired to a police officer. They were informed that they had to remember an unfamiliar face and later construct a composite with as much likeness as possible to their memory of this face. They were taken to a computer in a darkened and undisturbed environment and left, with headphones on, to watch the video. Recruiting and showing the video was done by a different person than the operator to ensure that the operator remained blind to the targets as well as to avoid providing physical context-reinstatement cues during retrieval.

Two days later the participants returned to the laboratory to construct the composites. This delay was chosen to achieve a sub-optimal memory for the target during the time of retrieval. Since police guidelines in the UK and SA define an maximum response time of 3 days, a 2-day delay was utilized in this study, as was done by Frowd et al. (2005b) and Schmidt and Frowd

(2006). Participants met with the operator individually in a separate room without outside interference. To control for possible differences in operator skills (Christie et al., 1981; Davies et al., 1983; Gibling & Bennett, 1994), an operator was trained extensively beforehand in both systems. This operator was blind to the target faces as well as to the study design. Participants were told they would make a composite at their own pace to give the best likeness, and encouraged to try and enhance the composite until they were satisfied with it or could not reach a better likeness. After the rapport-building stage and the outlining of the process of construction with this specific software, some participants underwent a context-reinstatement interview. The session then moved to the construction of the composite. Participants could select and change as often as they wanted to, with the operator staying with the participant and helping them apply the changes whenever requested. The session was ended when the participant stated that she had reached the best possible likeness. The sessions took between 0:45 and 1:20 hours, with an average of 0:55 hours.

Composite Evaluation Stage

Composites were evaluated using two separate methods. In the likeness rating, participants had to compare a composite with a photograph of the corresponding target. Likeness ratings might not be considered as having much forensic realism, but seem to be a more sensitive evaluation tool in research (Davies & Valentine, 2006). Composites were also evaluated with a from-memory line-up task, in which participants had to identify the target from a six-person-photo-array. Both evaluation methods were administered online, a practice that has become more common in recent times (Dandurand, Shultz, & Onishi, 2008; Heerwegh, 2006; O'Neil, Penrod, & Bornstein, 2003). A bespoke computer program was written to present the composites and targets, and record subject responses[4].

Participants

Participants for this stage were recruited from a different South African University, having a population similar to the participants from the composite

[4]The program was written by David Nunez, then based at the Department of Mathematics, University of Cape Town, as were all web-based programs used in subsequent studies.

construction stage. The participants had not seen the composites or target faces before, and were only allowed to take part in one of the two evaluation tasks. As an incentive, participants were entered into a lottery for an IPod. 508 participants took part in the line-up identification task, and 476 participants in the likeness rating task. The average age was 22 years (SD = 4.4 years), and 97% of the participants were male. 79% of the participants described themselves as White, 14% as Coloured, 6% as Black and the remaining 1% refused to give particulars[5].

Materials

Photographs were taken of each target, matching face position and lighting conditions to foil photographs. All target photographs as well as composites produced were standardised to 200x250 pixel images with roughly uniform interocular distance and white background.

For the line-up task, foils were selected from a large database of face photographs, taken under standardised conditions. Foils were matched to the description given by two participants not involved in the other parts of the experiment, using common characteristics as guideline for the choice of foils photographs. For each target two 6-person-line-ups were constructed (target-absent (TA) and target-present (TP)), with the same five foils being used in both line-ups.

An adjusted mock-witness scenario (Malpass, 1981) was utilized to determine line-up fairness and effective size. A good composite does not necessarily have to enable the identification of one specific person out of a group of very similar looking foils, but narrowing down the pool of possible suspects. A target can for example have a very specific hair style, and if the composite possesses this hairstyle it has to be considered a gain in information. Therefore, even though foils were matched to description, the description given to the mock witnesses was very broad ("young white male"), being true for all images. This was done

[5]In South Africa, where racial discrimination has only recently officially been abandoned, the issue of racial affiliation stays a sensitive topic. As such no participant was forced to give a definite answer, but was free to not choose at all. Prior to 1994 South African citizens were classified and registered as White, Coloured, Black or Indian according to the Population Registration Act of 1950. In this context, Coloured is differentiated from Black and refers to the group of people who are technically of mixed origins. These labels remain important in cultural or legal matters, e.g. Affirmative Action policies to reverse prior discrimination. Since racial classification is still a sensitive issue, no participant was forced to classify him-/herself as belonging to a specific racial group.

to provide a baseline for composite evaluation, and to ensure that the line-ups weren't biased due to one or more photographs standing out. Randomised sets of either TA- or TP-line-ups were given to 20 participants each. Effective size, calculated according to Tredoux (1998) was between 3.1 and 5.2, with an average of 4.0. The target or target-replacement for each line-up was not chosen above chance level (as indicated by the effective size of each lineup) (Malpass & Lindsay, 1999).

Procedure

Participants accessed the webpage through a link sent to them via email. In accessing this link from their computer they entered the first screen of the experiment. After being asked to provide demographic data, participant-judges received instructions.

Participants doing the likeness rating were instructed to look at both composite and target photograph, shown next to one another on the screen, and to decide to which degree the composite resembles the person in the photograph, that is, if they would have picked him from a crowd of people if they had to choose by using the composite. Likeness ratings of composite and target image were made on a 7-point-scale, where 1 = not at all similar and 7 = very similar. Each participant had to rate each of the 18 composites, with the order of composites presented being randomised.

Participants in the line-up task were instructed that the line-up may or may not contain a photograph of the target depicted in the composite. The composite was then presented to the participants for 20 seconds; a new screen then showed a line-up. The participants had to click on one of the photographs or on a button saying "none of them" to indicate their choice. Each participant had to identify one composite of each target, in total three composites. This approach led to a mixed design: participants had to identify each target only once, with the order of targets and TP/TA being randomised. This design led to a significant reduction of necessary participant numbers.

Both tasks were self-paced, containing no time limit for completion.

Results

As can be seen in Table 4.1, likeness ratings averaged over participants and targets were in the lower half of the scale, averaging around 2.25. A 2x3 repeated measures ANOVA revealed a significant effect for system, $F_{(2,950)} = 30.39$, $p < 0.01$, $\eta_p^2 = .05$ with no significant effect for interview or interaction. Although the effect size is small, Identikit as the featural composite system achieved consistently better ratings than ID.

Table 4.1. Likeness Ratings averaged by System and Interview Method

		Interview					
		Control		Featural		Holistic	
		Mean	SD	Mean	SD	Mean	SD
System	Identikit	2.71	1.01	2.37	1.05	2.77	1.08
	ID	2.00	.93	1.83	.82	1.80	.77

For the line-up task Table 4.2 summarizes the percentage of correct choices for each condition, that is, choice of the target in the target-present condition and rejection of the line-up in the target-absent condition.

Table 4.2. Correct Line-up Choices by System, Interview Method and Line-up Condition

		Interview					
	System	Control		Featural		Holistic	
		n	%	n	%	n	%
TP	Identikit	18	14.9	20	14.9	25	20.7
	ID	16	12.3	7	6.0	11	9.9
TA	Identikit	41	28.9	47	32.0	53	38.7
	ID	52	40.0	51	44.0	51	51.3

Hierarchical log-linear analysis (backward elimination method) was used to determine which model provided the best fit to the data. Even though, strictly speaking, a log-linear analysis might not be the best technique for partially repeated measures, it is considered a conservative enough measure to accommodate this. A 2x3 HILOG was performed to examine the effect on composite construction system and interview method on correct identification for both TP and TA- condition. In the TP-condition, there was no significant effect of interview method ($\chi^2_{(2)} = 2.59$, $p = .27$) or interaction between both independent variables (($\chi^2_{(2)}$ = 2.27, p = .32), but a significant effect of system ($\chi^2_{(1)}$ = 8.67, p = .00). Participants using Identikit as face construction system produced composites 1.7 times more likely to be recognized than did those using ID. In the TA-condition, both interview method ($\chi^2_{(2)}$ = 6.39, p = .04) and system ($\chi^2_{(1)}$ = 11.80, p < .00) showed a significant effect, with no interaction between both ($\chi^2_{(2)}$ = .00, p = .99). Post-hoc simple comparisons showed a significant difference between control condition and holistic interview ($\chi^2_{(1)}$ = 5.92, p = .02), and no significant difference when comparing featural interview to control ($\chi^2_{(1)}$ = .55, p = .47) and to holistic interview ($\chi^2_{(1)}$ = 2.84, p = .11).

Discussion

The results of this experiment are ambiguous in two respects: the advantage of the feature-based system Identikit over the holistic system ID on two quality measurements, but the opposite effect on the other quality measurement, and also the lack of an effect of the interview method. To be more precise, the holistic system showed an advantage when quality was measured with a recognition task in which the members of a target-absent line-up had to be compared to the memory of the facial composite and be correctly rejected as not displaying the target. The featural system on the other hand showed an advantage when the target was present in the line-up, and when quality was measured with a comparison task during which both photographed target face and facial composite were visible. These results would suggest that the correct rejection from memory is more related to holistic aspects of the face that are being promoted in composite construction with holistic software, and that identification of the target from memory and similarity of the composite in view of the target are supported by qualities related to more feature-based systems. The interested reader is referred to Frowd et al. (2004) and Tredoux et al. (2007) for a further discussion on the topic. Since context-reinstatement is the main focus of this chapter's studies I will focus on the interview method used. The benefits of systems that are based to a larger extent on holistic manipulation than featural manipulation have been described in detail elsewhere (Davies & Valentine, 2006).

In contrast to the effect of system on the measures of composite quality, the different context reinstatement strategies did not have a consistent effect on composite quality. One explanation for this result might be that the systems used provide the participant who constructs the composite with sufficient information on the central cue, thus not needing additional retrieval cues to remember the target face. The reason why context-reinstating interview methods had no consistent influence on composite quality might have been that face composite construction involves processes resembling recognition rather than recall. Witnesses might be able to recognize facial features resembling the target, and reject features that do not look alike. Conversely, Berman and Cutler (1998) found that even recognition of faces can be improved by asking the witness to attribute personality traits to the target face. Furthermore, Davies and Oldman (1999) found that personality attribution can influence later composite quality, suggesting that composite construction, even if it is not only a feature-by-feature recall exercise, can be influenced by changing the retrieval strategy.

Apart from the general influence of any manipulation on the compositing process, several moderating variables could have led to the ineffectiveness of context-reinstatement. Firstly, the chosen time interval between encoding of target face and composite construction might be a critical variable in determining the effectiveness of context-reinstatement techniques. As Malpass (1996) has pointed out, if the "focal" element of the event, in our case the target face, is readily available, additional context cues might not be needed, whereas if the memory for the target face is weak due to a longer delay, contextual reinstatement should be useful. In the domain of face recognition research, Shapiro and Penrod (1986) found that a decrease in recognition rate is approximately linear to the delay between encoding and retrieval of the face. Similarly, Deffenbacher (1986) found a highly reliable effect of delay on face recognition performance ($r = 0.25$). A more recent meta-analysis (Deffenbacher, Bornstein, McGorty, & Penrod, 2008) also found a strong negative effect of retention interval on memory for the once-seen face ($r = .18$). Even though some studies found that a deficit is not measurable for several weeks (e.g., Laughery, Fessler, Lenorovitz, & Yoblick, 1974), newer studies found a decrease in the amount of details witnesses could recall even after 24 hours (Ellis, Shepherd, & Davies, 1980). These studies lead to the conclusion that composites produced after two days are most likely worse in quality than those constructed only hours after the encoding of the target face. Considering the longer retention interval found in police practice (see Chapter 5), it stands to reason that the target face might be less readily retrievable, and thus context reinstatement more effective. Nevertheless, knowing that memory strength for faces might decrease over time does not yet allow a prediction about memory strength after any particular retention interval, as has been discussed in Chapter 2.

Secondly, the experimental manipulation between interview conditions might have been too weak for an effect to carry. The differences between conditions were evident in only a few questions asked. Also, all three interview conditions contained an initial rapport-building stage and questions related to the appearance of the target. This might have triggered the reinstatement of context in all participant-witnesses, with no further encouragement needed to recall the event and the target face.

Thirdly, and linked to the aforementioned reason, the encoded context might not have offered strong retrieval cues. Even though in other studies using video footage as context-reinstatement yielded an effect on the number of details recalled (e.g., Davis, McMahon, & Greenwood, 2005), the retrieval of faces might demand stronger contextual cues. Luu and Geiselman (1993) applied a Cognitive Interview before composite construction of a target face seen on video, and did not find an improvement in composite quality for all experiment conditions. On the other hand, Davies and Milne (1985) used a salient live encoding event, and found strong effects for both physical and mental reinstatement on composite quality. An additional advantage to a staged live event is that participant-witnesses would not immediately know that a face perception task will ensue, feeding into the demand for unintentional learning (see Chapter 2 and 3). Laughery, Duval, and Wogalter (1986) and Wells and Turtle (1988) found that participants tend to shift to a more feature-based encoding if they know that they have to later describe or draw a likeness of the target face. Based on these considerations, a second study was designed.

To be able to answer the question of the effectiveness of context reinstatement, a different study design was employed, adjusted to fit to the above mentioned considerations. In order to avoid similar effects, an interview technique was utilized that incorporated both featural and holistic approaches and should result in a more powerful mental context-reinstatement-interview. Utilizing another aspect of context-reinstatement, several studies found an increase in recognition and recall rate when witnesses were brought into the same environment during retrieval (physical reinstatement). Using both forms of context-reinstatement, the most effective experimental manipulation should be when witnesses are exposed to physical as well as mental context-reinstatement, thus not only having visual retrieval cues (external cues) but being encouraged to use their own storage context (like emotions, accompanying actions and perceptions) to encourage face retrieval. Also, as mentioned before, a live event in a rich context environment that is potentially emotionally charged might provide stronger contextual cues for later retrieval. Consequently, a live mock witness event was employed.

Past bad performance of face composites has been attributed to many factors, such as weak encoding, failure of the composite system, or a general yet unspecified incompatibility between task at hand and natural human face processing. In order to test encoding strength of the target, pilot witnesses, having had the same exposure than composite witnesses, underwent a recognition task. Composite witnesses' memory of the target face could naturally only be tested after the construction of the composite. Wells, Charman, and Olson (2005) found that witnesses who constructed a composite were significantly worse in a later recognition task than those who did not. In using potentially biased recognition performance from the composite witnesses we might not be able to establish encoding strength of the target face. On the other hand, Dekle (2006) and Meissner and Brigham (2001a) found no negative effect of composite production on recognition performance. My approach was rather conservative, controlling for this effect even though its existence is not verified.

4.3 Study 2

Summary of Experiment

A 2 (mental reinstatement) x 2 (physical reinstatement) between-subject design was employed. A retention interval of 16-18 days was exerted, emulating the average delay observed in a study of South African police (Schmidt & Tredoux, 2006). Testing encoding strength and memory for the target, a first group of participant-witnesses who constructed the composite and a second group of participant-witnesses who did not construct composites were later tested for their recognition performance. As composite evaluation methods, the same methods as in Study 1 were employed, except that the identification task was done only with target-present line-ups due to restrictions in the number of participants who could participate in this live event. Composites can be found in Appendix E-06, and photo-arrays can be found in Appendix E-07.

Composite Construction Phase

Participants

108 participants were recruited at the University of Cape Town by means of an advertisement, and paid ZAR 20 (2.60 US$) for taking part in this study. They were offered further information on the outcome of this study if they were interested. No exclusions were made on the basis of race, but participants who knew the target person they encountered would have been excluded. For each event, eight participants were to be included in the experiments, amounting to a total of 96 participants. The average age was 21.3 years (SD = 3.9), and 73% participants were female. Participants were randomly assigned to two groups, one which constructed composites and one which did not.

Materials

Twelve young White males were chosen as targets. These mock perpetrators were recruited from the drama department at the University of Cape Town by advertisement, and chosen according to the same requirements outlined in Study 1. All of them had previous acting experience, and were not usually on the main campus of the university.

For each taret, a six-person line-up was constructed in the same way and with the same fairness measures as in Study 1. Effective size, calculated again according to Tredoux (1998) was between 3.1 and 5.4, with an average of 4.5. As in the previous study, no target was chosen above chance level. In contrast to Study 1, only target-present arrays were used to keep the number of participant-witnesses within a manageable range for the mock event.

All staged events followed a given script, which had been written by the experimenter and was repeatedly rehearsed with each mock perpetrator beforehand. Targets and experimenter followed this detailed script as closely as possible in order to achieve comparable events, and to provide a common basis for the interview administered during face retrieval. The script can be found in Appendix P-01. Participants were present in the room, ostensibly recruited to take part in what they believed to be a short-term memory test. During this test, before as well as after the staged event, participants were required to focus their attention on the experimenter due to the short-term memory test requirements. They were seated in a half-circle around the experimenter and opposite the door when a knock at the door disturbed the ongoing mock test. A person entered the room and demanded several items to be handed to him that he claimed had to be in this particular office. He then commenced an angry argument with the experimenter, and asked participants several questions. To ensure that each participant paid attention and had the chance to encode his face, the intruder made repeated eye contact with each participant and, searching for the items in the room, came at least as close as about one meter to them. Due to these precautions, only a limited number of participants could take part in each event. He then left the room, and the mock memory test continued for another five minutes before the experimenter revealed the true purpose of this event. Staged events were timed, and took between 70 and 90 seconds.

The software system Identikit 2000 was chosen in combination with a graphical software tool (Photo-Shop CS), which was used to enhance the composites afterwards as described in Study 1. Because the number of participant was limited due to the study design, only the software system currently utilized by the South African Police Service (see Chapter 5) was used.

Procedure

All participants were initially asked to come to the department for a short-term memory test, and to return 16-18 days later for a second testing. At the first appointment they came to the department and were seated in an office together with seven other participants. The office, by its nature, contained a lot of objects which could later be used as contextual cues when participants returned to the same office in the physical-reinstatement condition. The experimenter then explained the task at hand, a short-term memory test in which they had to remember numbers given by the experimenter either verbally or visually, and to write them down after the stimuli were removed. The results of this test were of no further relevance for this experiment, but were given to the participants if they were interested. During the visual encoding task, the prepared mock crime took place as described earlier. After the target had left, any discussion among participants or with the experimenter on the topic was impeded, and the original test was resumed. Five minutes later the experimenter revealed the real purpose of this meeting, and asked the participants to imagine them having been eyewitnesses to a crime, and to aid the police in the second appointment with their memory of the target. After giving their informed consent to participate, they were asked to return to individual appointments.

At the second appointment participants were led to either the same room or a different bare room in another part of the building. They were then greeted by the operator, without seeing the experimenter from the first appointment again. The operator was both trained in interviewing and Identikit by self-study, coaching from an experienced interviewer, on-the-job training by an experienced police officer as well as several trial runs prior to the experiment. This operator was not involved in prior stages of this experiment nor the prior experiment, and was blind to the target's appearance, number of targets, as well as which target each participant had seen. He informed the participant about the procedure that was going to be used. After a rapport-building stage he then either used the mental context-reinstatement techniques to aid the participant in remembering, or went straight to the retrieval task.

Participants in the mental context reinstatement group were asked to remember when they came to the department to take part in the mock test. They were then asked to recall the staged event in detail, starting this retrieval with questions about the physical environment, their activities, thoughts and emotions before, during and after this event. They were then asked to remember the physical appearance of the target as well as their associations and impressions, without verbally answering questions regarding the face. The latter instruction was given to avoid the possible verbal overshadowing effect as discussed earlier in this chapter. Participants were encouraged to take their time. When no more information could be elicited and the participant described the target face as being clear in their mind, the operator asked them to once again picture the face, and then to continue on with the next task (composite construction or target identification). The mental context reinstatement interview was implemented to improve retrieval of the target face for subsequent composite construction and/or recognition task.

Half of the participants were asked to construct a composite of the target, using the same program and techniques as described in Study 1. After the composite was finalized, participants were asked to identify the target from a six-person-array. Participants who did not have to construct a composite were asked to identify the target from the same six-person-array in order to estimate encoding strength of the stimuli. All participants were subsequently debriefed and dismissed. This second appointment took between 10-15 minutes for recognition-only participants, and 40-55 minutes for participants constructing a composite.

Composite Evaluation Stage

Composites were evaluated using two separate methods, as described in Study 1. In contrast to Study 1 only target-present line-ups were used in the identification task due to the limited number of available participants for each target. Composite quality was determined by simple identification rate, that is, a higher identification rate of one specific composite was interpreted as higher composite quality. In other words, rejections rates in target-absent line-ups were not available.

Participants

Participants' inclusion and administration followed the same procedure as in of Study 1. 1084 participants took part in the line-up identification task. 43 participants took part in the likeness rating task. The average age was 22 years (SD = 6.6 years), and 86% of the participants were male. 61% of the participants described themselves as White, 15% as Coloured, 22% as Black and the remaining 2% refused to give particulars.

Results

Separate analyses were conducted for composite identification and likeness rating. Prior to this, influence of composite construction and encoding strength were tested.

Encoding Strength

A total of 96 participants that were included in this experiment saw one of the targets in the staged event. Of these, 48 participants subsequently constructed a composite. All participants then attempted to identify the target in the target-present line-up. These groups did not differ in terms of age ($t_{(1,94)} =$ -.41, p = .681) or gender (($\chi^2_{(1)}$ = .84, p = .491). Recognition accuracy was 68%. In the recognition-only condition 34 participants correctly identified the target as opposed to 31 participants in the composite-construction-condition. Participants who had constructed a composite did identify the target equally well to the recognition-only participants in the later recognition task (($\chi^2_{(1)}$ = .43, p = .663). Although probing the effect of composite construction on later identification performance was not the focus of this study, results show that composite construction had no effect on the later recognition of the target from a line-up. The recognition rate shows that participants remember the target face sufficiently well to recognize the target. Consequently they should have had a reasonably good memory of the target on which the composite is based. On the other hand, these numbers can be used to estimate the difficulty of the line-ups for composite quality evaluation.

Likeness Ratings

As was the case in Study 1, likeness ratings averaged over participants and targets were in the lower half of the scale, with M = 2.8 (SD = .91). The data was subjected to a 2 (interview type) x 2 (physical reinstatement) repeated measures ANOVA. This produced a significant main effect for mental reinstatement, $F_{(1,42)} = 21.25$, $p < 0.01$, $\eta_p^2 = .34$, indicating a detrimental effect of mental reinstatement on composite likeness, and no effect of physical reinstatement, $F_{(1,42)} = .048$, p=.827, $\eta_p^2 = ..001$. The interaction between these factors was significant, $F_{(1,42)} = 14.05$, $p < 0.01$, $\eta_p^2 = .25$, with mental context-reinstatement having a stronger influence on composite likeness when physical context was not reinstated. See Table 4.3 for details.

Post-hoc simple comparisons revealed a significant decrease in rated likeness through mental reinstatement when composites were constructed in a different room ($F_{(1,42)} = 37.58$, $p < 0.01$, $\eta_p^2 = .47$), but no significant difference when in the same room ($F_{(1,42)} = 2.17$, $p < 0.01$, $\eta_p^2 = .05$).

Table 4.3. Likeness Ratings averaged by Physical Reinstatement Method and Interview Method

		Interview Method			
		No mental reinstatement		Mental reinstatement	
		Mean	SD	Mean	SD
Physical Reinstatement	Different room	3.02	.932	2.55	.794
	Same room	2.83	.953	2.71	.945

Identification Accuracy

For the line-up task Table 4.4 summarizes the percentage of correct hits in each condition, that is, how often the target face was chosen on the basis of the composite face. Since only target-present line-ups were used, hits are defined as correct recognition of the target face.

Table 4.4. Correct Identification by Physical Reinstatement Method and Interview Method

		Interview Method			
		No mental reinstatement		Mental reinstatement	
		n	%	n	%
Physical Reinstatement	Different room	69	25.1	80	28.4
	Same room	63	24.0	73	27.8

Hierarchical log-linear analysis was used as described in Study 1. A 2 x 2 analysis revealed no significant effect for mental ($\chi^2_{(1)}$ = .35, p = .554) or physical reinstatement ($\chi^2_{(1)}$ = .24, p = .628), nor for the interaction between both independent variables ($\chi^2_{(1)}$ = .01, p = .926).

Three measures of identification levels have been presented so far, all utilizing the same line-up. First, effective size of the line-up was measured, putting the adjusted chance level of participants identifying the target at 22.2%. Second, recognition of the participant-witnesses who had seen the target person was measured. Only utilizing rates from the recognition-only participants, participants identified the target in the line-up in 70.8% of all cases. Thirdly, participants who only saw the composite identified the target, on average, in 26.3% of all cases. Participants who based identification on the composite identified the target above the adjusted chance level ($\chi^2_{(1)}$ = 10.33, p < .01). Participants were 1.2 times more likely to identify the target when basing their decision on the composite.

4.4 Discussion and General Discussion

Results led to the conclusion that both composite quality measures showed better than floor-level results: significant differences between conditions were found in the likeness rating measure, and identification rates were above adjusted chance level in the line-up identification measure. However, likeness ratings of the composites produced with or without mental context-reinstatement resulted in the only significant outcome: A decrease in likeness when participant-witnesses were asked to recall aspects of the event in question. Composite quality as evaluated by likeness rating and line-up identification in Study 1 and line-up identification in Study 2 showed no improvement through mental reinstatement of the original context.

Identification rates in the line-up task also showed that composites were not identified as often in a line-up as when the identification was based on real-life experience with the target person. This result was to be expected and in line with considerations regarding the retrieval of target memory for composite construction as discussed in Chapter 2. Furthermore, even using a photograph of a target face instead of a live event decreases the later identification accuracy of the target face considerably (Megreya & Burton, 2008).

The lack of consistent significant improvement of composite quality by means of context-reinstatement measurements suggests that if there is a positive effect of context-reinstatement, it can only be achieved under very specific conditions. In the experiments presented in this chapter, a holistic interview resulted in composites being more accurately rejected in a target-absent line-up than when composites were constructed without reinstating the context. Taking up the discussion in Chapter 3 about differences in the definition of composite quality, an improved rejection of line-ups could mean that composites constructed with the help of a holistic interview are more distinguishable, or that participants are more likely to reject the line-up in general, regardless of the similarity or distinguishability of composites and line-up members. If composites are more 'rejectable' when constructed with the help of a holistic interview, this might also explain results of Study 2. Participants here rated the likeness between composites constructed with the help of mental context-reinstatement and target faces lower than when constructed without the help of a context-reinstatement interview. Here again, participants might be led to have a decreased willingness to see similarity between composite and photographed face, thus rating the likeness lower. Unfortunately, Study 2 did not contain a target-absent condition, thus making it impossible to see if results from Study 1 could be replicated.

Other authors investigating context reinstatement and composite construction have found results different to those in the experiments described above. Davies and Milne (1985) found an effect of mental as well as physical reinstatement on matching accuracy. An important difference between their study and studies in this chapter is the use of additional information when evaluating composite quality. Participant-judges in the matching-task were provided not only with the composite, but also with “relevant portions of the subject’s verbal description” (Davies & Milne, 1985, p. 215). One could argue that composite quality between mental context-reinstatement conditions didn’t differ, but the quality and/or quantity of description provided for the comparison did. A good description of relevant features might facilitate feature-to-feature comparison in a matching task. A specific ponytail, especially with female targets, might be easy enough to describe, and might prove to be a better match to the target than any ponytail provided by the system. Great care has to be taken not to confuse the influence that context-reinstatement can have on composite quality with the influence mental context-reinstatement has on the description of the target.

In contrast to the studies in this chapter that found no consistent advantage of context reinstatement, Frowd et al. (2008) found that adding free and cued recall of personality traits of the target to the pre-construction interview improved composite naming, sorting and likeness ratings. Participants were asked to watch a video with the intended purpose of later composite construction, an approach that seems to benefit a more featural encoding of the target face (Wells & Hryciw, 1984; Wells & Turtle, 1988). After 3-4 hours participants underwent a procedure labelled Cognitive Interview[6], in which participants were

[6]The technique labelled ‘Cognitive Interview’ in this study was an adapted context-reinstament interview form by the British Home Office, mentioned earlier, containing visual imagery and recall as well as cued recall of featural information.

asked to describe and rate facial features. After this feature-based retrieval task, some of the participants were asked to describe and rate seven personality-traits associated with the target face in the video, a procedure Frowd et al. (2007a) called 'Holistic Interview'. Frowd et al. (2008) found effect sizes between $\eta_p^2 = .55$ and $\eta_p^2 = .8$ for the different evaluation tasks. In an earlier study Frowd et al. (2007a) compared the UK-version of the Cognitive Interview with only the "Holistic Interview", finding no significant effect of the former interview. In a study utilizing a series of personality judgements as holistic context reinstatement, comparing those to an adapted Cognitive Interview (see above), Frowd et al. (2005a) found an effect of Cognitive Interview over no interview, but no difference between the holistic interview and the other two conditions. It will be interesting to see if the reported advantage of the added 'Holistic Interview' in Frowd et al. (2008) can be replicated.

Chapter 5

Utilisation and Usefulness of Face Composites in the South African Police Service

This chapter describes two field studies of face composite practice within the South African Police Service. The studies were conducted over a 30-month period, and describe police practice in cases in which face composites are constructed. These studies analysed police practice of face composite construction to address several pertinent issues in the literature of composite construction and eyewitness memory.

5.1 Background and Aim

The aim of this research is to gain knowledge of the conditions under which composites are constructed in police practice to draw comparisons between real-life conditions and conditions in a laboratory setting. Composite researchers often advise on best-practice to be followed by the police. As will be argued in this chapter, it might not be effective to advise on best-practice before real-life conditions are investigated. It is for example unclear how often a composite is needed. Moreover, an investigation of police practice can show how frequently investigations can advance significantly without utilizing composites (McNamara, 2009; Wells et al., 2005). Investigating police practice can also show in how many cases more than one witness can construct a composite (Bruce et al., 2002), and what the average delay between crime and composite construction is (Davies et al., 1978a, 1978b; Green & Geiselman, 1989; McNeil et al., 1987). This chapter will first elaborate on the rationale of field studies, their benefits and limitations, before introducing ideas on how field studies in the domain of composite research can be designed and implemented.

Field Research – Merits and Demerits

One of the chief postulates scientific research has to answer to is the criterion of replicability (Malpass et al., 2008). A diverse range of studies has to reach similar conclusions regarding a specific phenomenon before a finding can be promoted, implemented, defended as evidence through expert testimony, and used to advise police to change practice. Diversity in this context means that a finding has to be replicated over a range of laboratories and study conditions. In the case of composite construction, the majority of research has been done in the laboratory, with most studies adhering to a rather narrow range of design factors such as a short delay between exposure to the stimuli and composite construction (see Chapter 3). To establish composite quality and the enhancement thereof through manipulation of certain factors, a wider range of research approaches is needed.

Best practice in composite construction can currently be derived only from experimental studies using either staged incidents or, more commonly, a video or photo presentation. These recommendations of best practice, if given, do not take into account the lack of external validity inherent in this field of research, and which has to be addressed (Turtle, Read, Lindsay, & Brimacombe, 2008; Yuille, 1993). Behrman and Davey (2001) and Steblay (2008) advocate increased studying of real crimes and testing results in real-life settings to establish that changes in real-life investigations will be beneficial. Field studies are a workable and necessary step to achieve best-practice recommendations for face composite construction. In other fields of eyewitness testimony, this approach is already being implemented, with changes in policy on police procedures having been made both in the United Kingdom (e.g., Devlin, 1976, cited in Wells et al., 2000, p.592) and the United States of America (e.g., Wells et al, 1998).

Investigating real crimes allows for the study of a wide range of variables usually not found under laboratory conditions. Yet the variation of many different variables in real life makes it more difficult to deduct cause-and-effect chains from inferential results. Field studies cannot usually exert the same control over study conditions that laboratory experiments can. To establish causal relationships between variables in the field is therefore an aim difficult to achieve. Consequently field studies should not necessarily be judged according to the same standards that laboratory studies have to adhere to. Investigations of real crimes should be seen as a different research approach, aimed at evaluating practice rather than finding causal relationships.

Field Research in the Area of Eyewitness Testimony

In addition to laboratory experiments, two main research approaches have been defined by Wright, Boyd and Tredoux (2001) in the field of eyewitness testimony: The use of archival data, and conducting of field experiments. Archival studies access information from real-life police cases. For example, Behrman and Davey (2001) used police reports to assess line-up practice, a research approach that utilizes data already reported in one form or another. Although this approach does not intrude on any current investigation or alter reports by using data that has already been reported, the researcher has to be satisfied with existing data already in the police docket. This might lead to a study not being able to collect all variables required. For example Behrman and Davey were unable to analyse suspect and foil identification in all studies, as only one police station collected data on foil identification.

A crucial point discussed in archival studies based on eyewitness testimony is the differentiation between suspect and perpetrator. Unlike in a simulated crime in a laboratory, suspects might be either guilty or innocent. Not even convictions are proof of guilt, as has been impressively proven by the Innocence Project and other institutions, who lobby for DNA-exonerations of convicted innocents (e.g., Wells, 2008). The judgement 'guilt' or 'innocence' is not reached independently of eyewitness testimony. One exception is the state of California, where eyewitness evidence is not admissible in court. In these cases only other extrinsic incriminating evidence is used to determine "guilt", resulting in a more independent criterion for field studies (Behrman & Davey, 2001; Behrman & Richards, 2005). Although not independent of identification evidence, one could nevertheless argue to use the criteria for "guilt" as the legal system does, defining a conviction as best possible evidence that the suspect is indeed the perpetrator being depicted in the composite.

Another form of collecting data from real-life cases is the use of survey questionnaires. Using a questionnaire form does not restrict the research to data that can be found in police files, but can assess all variables the researcher deems important. A questionnaire form can also assess police practice by asking police officers about their experience regarding police practice. Kitson et al. (1978), using a questionnaire, evaluated the extent to which the first version of Photofit was utilised within the British police service to advise manufacturer on necessary alterations to the kit as well as to advise police users about techniques related to the kit. They asked police forces to report every case in which a composite was made within a 6-month period, and report information on the construction process as well as on composite use. McQuiston-Surrett et al. (2006) also used a survey approach to evaluate police experience regarding composite construction, by sending a questionnaire to law enforcement agencies.

The second approach within the area of field research is conducting experiments in a natural setting (Wright et al., 2001). Field experiments are considered a compromise between external and internal validity. Their advantage over archival studies lies in the possible assignment of participants to experimental groups, and in that participants are not immediately aware that they are part of an experiment, thus giving the impression of being a real-life situation. However, field experiments face different challenges, like attrition or reduced experimental control. Since no field experiment has been published in the area of composite construction or will be discussed in this thesis, I refer the interested reader to Valentine and Messout (2009) and Wright et al. (2001) for examples and a discussion on field experiments within the area of eyewitness recognition.

Field Studies on Composite Construction

A rare investigation into police use of composites is a survey done by the British Home Office (Kitson et al., 1978). This study was done five years after the introduction of Photofit, and was aimed at enhancing the kit as well as police training regarding its use. Following up on over 700 cases in which composites were constructed in 1976, investigating officers reported that the investigation benefited in one out of four cases in which a composite was constructed. Officers also reported that only 10% of composites were published in the media, while the majority of the composites were shown to informants and to people in the community where the crime had happened or to individuals who were in some way involved in the crime. Investigating officers also stated that in 25% of all closed cases, Photofit had played a key role in solving the case.

Nearly 30 years later, McQuiston-Surrett et al. (2006) conducted a survey concerning composite practice in US law enforcement agencies. Although more than 1600 questionnaires were distributed, only 10% were returned. According to answers given by the police officers responsible for composite construction, nearly all composites were constructed with computerized systems, Identikit 2000 and Faces 3.0 being the most frequently used among them. No standards were found among composite officers about the procedure with which the witnesses were interviewed prior to, or during composite construction, which the authors considered a shortcoming that should be addressed, and which has been addressed in the previous chapter.

What is noticeably absent in published research is objective data on the importance of composites in terms of investigative proceedings. Publications over the last 30 years comment on police practice by referring to personal experience or communication with an experienced police officer (e.g., Bennett, 1986; Boylan, 2000; Taylor, 2001). Frowd et al. (2005a, 2005b) state that composites are published by the police and recognized by members of the public as persons they are familiar with. Without an investigation into how composites are predominantly used, and if it is mostly based on the recognition of familiar persons, this utilisation of composites is only one out of many possible utilisations. As discussed in Chapter 3, there are many possible functions of composites, all of which have implications for the measurement of composite quality in experiments. I therefore consider it important to move composite research towards a better understanding of how research can be meaningful to police practice by studying composite construction in real-life. Thus, the following studies are aimed at demonstrating some of the problems in actual cases involving face composites, as well as to be able to compare experimental settings with real-life conditions for composite construction.

Composite quality depends on many factors. Some of those factors have already been acknowledged for their influence, most prominent of all being the system with which the composites are constructed. Composite quality seems to correlate negatively with time between crime and construction, when a witness belongs to a different race or age group than the suspect, and with perceptual opportunity (Davies & Valentine, 2006). Other factors, as described in Chapter 3, were studied in only but a few experiments, and it remains to be seen if they yield a permanent influence on those cognitive processes involved in the construction process. In the field of eyewitness identification, several of the eyewitness's memory aspects that also pertain to memory of faces for composite construction have been examined in great detail. Behrman and Davey (2001) specifically highlight the importance of delay, cross versus own-race, and weapon focus effect in determining memory and subsequent identification accuracy. The identification of a member of a different racial group is another factor influencing eyewitness testimony in laboratory studies (Meissner & Brigham, 2001b) and field experiments (e.g., Wright et al., 2001), and might therefore be an important factor in composite construction in South Africa due to its racial diversity. However, the only way to determine factors influencing the quality of composites constructed under real-life conditions is to evaluate composites constructed under real-life conditions.

5.2 Study 3 - Archival Study of Composite Practice in the South African Police Service

Methodology

Studying Composite Practice within the South African Police

A specialized unit of the South African Police Service (SAPS), the Local Crime and Record Centre (LCRC), deals with forensic evidence and constructs composites in South Africa. Any LCRC serves a bigger area than any single police station. The LCRC Cape Town, for example, serves police stations up to 80 km from Cape Town centre. Here, specialized police officers compile composites on a computer screen according to instructions given to them by the witness(es) after having undergone the general police training and a training course for constructing composites. These composites are then given to the investigating officer, based in the detective units of each police station branch, to be used in the investigation.

Permission to do research on composite construction was granted by Commander De Beer of the Police Headquarter in Pretoria in 2004. Superintendent Van der Westhuizen, then head of the composite unit within the LCRC Cape Town, facilitated access to information from the LCRC and access to detective units.

Sample

All cases in which composites were constructed by the LCRC Cape Town in 2002 were included in the sample. This year was chosen to increase the likelihood that the majority of cases had been settled in court, and that, if possible, convictions would have been reached. In total 520 cases in which one or more composites were constructed were used for further analysis. Please note that as a rule only one composite per suspect was constructed, even when more than one witness was able to construct a composite.

Survey Forms

A survey form for the LCRC was constructed, including questions about crime type, number of witnesses and composites, and the delay between crime, commission of the composite by the detective's unit, and compilation date. A second form was constructed for the detective's unit, requesting information on the case status. For cases in which a conviction had been reached, further information on the usefulness of the composite and the availability of other evidence would have been requested with a third form[7]. The first two forms can be found in Appendix P-02 and Appendix P-03.

Procedure

Data collection began in May 2005. LCRC data was collected first and in person. All cases were then clustered per police station, and the second form was sent out with a letter requesting data on those cases sampled from the

[7] A conviction was seen as the strongest indicator available available that the suspect is in fact the perpetrator of the crime in question, although, as stated before, it is by no means certain that this is indeed the case. The composite in these cases can be assumed to depict the convicted suspect, and as such should have a likeness. The quality of the composite, constructed in a real-life police investigation, can be assessed with means found in an experimental setting, as is described in Chapter 3.

LCRC dockets. In total, survey lists were sent to 74 different police stations. For cases in which a conviction had been reached, a standardised photograph of the convicted suspect should have been in the police docket, as well as information on the use of the composite. In those cases the third survey form would have been sent out, requesting information on composite utility and a copy of the suspect photograph. Composite quality in real-life settings could have been evaluated in a subsequent laboratory study, comparing the composite and photograph of the convict on which the composite should have been based. Data collection was finalized in August 2005, with a response rate of 100% for the first two forms. Utilization of the third form and a subsequent comparison of photographed convict and composite did not take place.

Results

Results are reported in descriptive manner only. An inferential comparison of the photograph of the convicted suspect and the composite could not be achieved due to a lack of both convictions and photographs, as will be discussed later.

The types of crime in which face composites were requested can be viewed in Table 5.1. The information was clustered according to broader categories than were found in the Record Centre files. In cases where two complaints were made (for example abduction and murder), the more severe category in terms of legal sentencing was used to categorize the case. Nearly half of the cases in which a composite was produced were armed robberies, followed by simple robbery and rape/sodomy.

Three dates were gathered in the Record Centre files: date of the crime, date of requesting the compilation from the Centre, and the date when a composite was compiled. Since delay between receiving the request and responding to it is more an administrative than a memory issue, Table 5.2 reports only the time delay between the witness encountering the crime and reconstructing the suspects' face.

The median time interval between crime and composite construction was 18 days. Most composites were constructed three days after the crime was reported.

Table 5.1. Type of Crime (clustered)

	Frequency	%
Armed robbery	259	49.8
Robbery	92	17.7
Rape / sodomy	57	11.0
Theft / corruption / fraud	50	9.6
Murder / manslaughter	27	5.2
Assault	16	3.1
Attempted murder	12	2.3
Abduction	7	1.3
Total	520	100

The average delay was much higher since some composites were constructed more than three years after the crime took place.

Table 5.2. Time Interval between Crime and Face Compilation

N valid	520
Mean (days)	33.93
Median (days)	18.00
Std. Deviation (days)	72.87
Minimum (days)	0
Maximum (days)	127

Face composites were done with one or more witnesses attending the compilation session, compiling one or more composites of suspects. Table 5.3 reports the number of witnesses who constructed the composite(s), and Table 5.4 the number of composites per case. The number of composites constructed does not necessarily relate to the number of suspects involved, since only the faces of suspects who could be remembered well enough were compiled.

The majority of composites were constructed by only one witness, with more than one witness participating in the construction in only 17.3% of cases. Similarly, in two out of three cases only one composite was constructed. According to the officers constructing those composites, witnesses mostly felt competent enough to remember the most dominant or demanding perpetrator, but not other perpetrators.

Of all the data originally intended to be collected from police files and the

Table 5.3. Number of Witnesses per Case

Number of witnesses		Frequency	%
Valid	1	429	82.7
	2	58	11.2
	3	21	4.0
	4	6	1.2
Total		514	98.8
Missing		6	1.2
Total		520	100

Table 5.4. Number of Composites per Case

Number of composites		Frequency	%
Valid	1	349	67.2
	2	134	25.8
	3	31	6.0
	4	4	0.8
Total		518	99.6
Missing		2	0.4
Total		520	100

administrative computer system only information on the legal status of each case could be acquired reliably for all cases. Of the 520 cases included, 45 were still open (either in court or under investigation) at the time of analysis. In two cases the information wasn't accessible. The remaining 473 cases were closed, as can be seen in Table 5.5. Of the 473 closed dockets, only four were closed with a conviction, the remaining 469 were closed undetected or withdrawn.

Data on the distribution of the face composites and other evidence was not consistently available in all the files. In personal conversation with several investigating officers it was repeatedly remarked that the utilisation of the composites would only be mentioned in either the witness statement or the investigating officer's report when the distribution of a composite leads to new evidence that can be used in the investigation. In the four cases where a conviction was reached, the researcher retrieved all available data on the use and usefulness of the composite and the suspect's photograph from the police stations. No information on composite use was found in the police files, and only one photograph of the suspect could be retrieved. Two police dockets did not contain a photograph, and one docket could not be found at all.

Table 5.5. Legal Status of Cases

Case status		Frequency	Percent
Valid	open	45	8.6
	closed undetected / withdrawn	496	90.2
	closed with conviction	4	0.8
Total		518	99.6
Missing		2	0.4
Total		520	100

Discussion of Results

Utilizing the available information from the LCRC brought interesting aspects of composite use and utility in the South African police to light. These will be examined briefly, followed by a thorough discussion on the shortcomings of this research approach, including recommendations for a second field study. A detailed discussion can be found at the end of this chapter.

Concluding from the LCRC files, time delay between the crime and compilation of the face composites was considerably longer in practice than in most laboratory studies, even when considering only the median value of 18 days. This long delay might have an impact on memory for the face and consequently composite quality. For a second study, it is important to take note that several factors have changed since 2002. In 2003, a policy paper was released by the SAPS (South African Police Service, 2003), stating that a composite should be requested from the investigating officer within 24 hours after the incident happened, acknowledging that the actual construction within the Centre might take much longer due to various administrative reasons such as limited resources and availability of the witness. Also, since 2005 the face composite unit within the LCRC was put under direct command of a superintendent whose task it is to manage the unit's activities as well as promote a more appropriate and timely use of its service. Both factors might have contributed to a change in administrative demands, and subsequently a change in delay between crime and composite construction.

Summarizing the type of crimes that were committed, one can see that the majority were crimes in which violence, physical harm, or at least the threat thereof, was present. This could, following critical consideration of laboratory-based research, have resulted in a different quality of event memory. Looking at photos of faces and then being asked to retrieve this information is in many ways not the same as being a victim of a rape or an attempted murder (Deffenbacher et al., 2004). On the other hand, the current research design does not allow the deduction that the witness constructing the composite was the victim of the crime. Neither can it be said that, for example in an armed robbery, a gun was used and pointed at the witness. Since both factors might have an impact on eyewitness memory, a second study will address this issue.

An interesting finding is that in one out of five cases more than one witness helped in constructing the composite. To date, only two studies have investigated the quality of individual composites and joined computer-merged composites (Brace et al, 2006b; Bruce et al., 2002), and no study has specifically investigated inhibition and/or support processes between witnesses in the composite construction process. For current police practice, this seems to be a minor concern since administrative considerations do not allow for multiple composites of the same suspect to be processed, and no time is allocated for separate interviews of witnesses having seen the same suspect, making a separate composite construction next to impossible (Supt. Van der Westhuizen, personal communication, October 10th, 2005). Nevertheless it is necessary to have knowledge about these processes in order to judge the outcome appropriately, and to see if the above mentioned policies are wise to pursue.

Regrettably it must be said that not all the objectives of the study could be fulfilled using this specific study design. The use of face composites once they left the Centre could not be assessed, nor can any conclusion be drawn about the quality of the face composite. Information could be obtained regarding crime and construction process, but answering the question of the effectiveness of face composites in the investigation would require the choice of a different research design. As was utilised in the study by Behrman and Davey (2001), it might be necessary to collect data during investigations in a more active manner, thus not having to rely on data available in the archives. In this study, data on composite use within the investigation was not available, giving no indication of whether the composite was used at all, and if so, whether it contributed to the investigation.

Using a conviction as criterion for 'guilt' might pose a problem in evaluating composite practice, at least in a country like South Africa with a low conviction rate in contact crimes. For this study, a conviction was reached in less than 1% of all cases. In contrast to this, a study investigating cases utilizing eyewitness testimony of any kind in Israel by Levi and Almog (1996, as cited in Levi, 1998, p. 395) found a conviction rate in eyewitness cases of 21%. Further studies would have to determine whether the conviction rate changed in subsequent years, or, if a criterion for 'guilt' was needed, to consider other criterion estimates such as other forensic evidence against the suspect (Halford, Milne & Bull, 2005).

In response to shortcomings of the first field study (Study 3), and to evaluate a broader range of variables potentially influencing composite quality, a second field study (Study 4) was conducted, employing a different study design. As was the aim of the first study, the second study evaluated use and usefulness of face composites. More precisely, this study aimed to find in which cases composites would be constructed, and if these cases depended on composites to further the investigation. It might not be possible to evaluate composite quality directly by comparing convict photographs with composites, as was the case in the prior study. However, as is done in the field of eyewitness identification, it is possible to approximate possible factors impeding witness' memory and subsequently composite quality. For example, viewing conditions and the involvement of the witness have shown to be factors influencing later eyewitness identification (Shapiro & Penrod, 1986), which might also be influencing other face retrieval tasks. This study also gives evidence on the utilisation of composites by the investigating officers. If for example composites are mainly used in isolation, and shown to members of the community in which the crime happened in the hope that someone will recognize an acquaintance, the naming task as described in Chapter 3 might be the most appropriate estimate of the usefulness of composites. The evaluation of these factors should draw a coherent picture of the current construction and use of face composites, and possible ways to improve both practice and future research.

5.3 Study 4 – Follow-up Study on Composite Practice in the South African Police Service

Methodology

Sample

Between October 2005 and November 2006 all cases in which composites were constructed by the LCRC Cape Town were eligible for inclusion into this study. Due to the relocation of the Identikit-unit from local to provincial LCRC in the last months of 2005, other administrative matters increased and the continuation of data collection was put on hold for several weeks. Also, only cases in which the police officer constructing the composite had at least several months of experience in this task were included. Seven police officers constructed an average of 73 composites each. In total 578 cases were included with a total of 815 composites constructed. Between one and four composites were constructed per case. 368 investigating officers at 119 detective units were asked to fill out the IO form (see survey forms). Because only 311 IO forms were returned, subsequent description of IO form-data will be based on only these cases instead of the total of 578 cases included in this study.

Survey Forms

This archival study used several different forms to gather information about witness, crime and suspect particulars, composite construction and use. Two questionnaires were constructed, with additional data collected from the electronic database at the LCRC and police files. In addition, three local newspapers were searched for published composites as independent verification of composite use. The latter attempt had to be abandoned since more often than not, composites were published without the related case number, which made relating composite to case very difficult.

Two questionnaires were constructed for this study. One was to be filled out by the Identikit member during or after each appointment in which a composite was constructed with a witness to a crime. This form, subsequently called 'LCRC form', contained questions regarding particulars of the crime, the witness and the session such as length of session or race of witness.

The Identikit member then attached the second questionnaire, subsequently called 'IO form' to the composite constructed. This form was intended to be filled out by the investigating officer and returned no later than three months after receipt of the form and composite. The form was divided into three parts, containing questions on utilisation of composite, legal status of the case and other evidence available. The LCRC form and the IO form are attached to this thesis as Appendix P-04 and Appendix P-05 respectively. Both questionnaires were designed such that it was as easy and quick as possible to fill them out. In addition to these two forms for each individual case, further supplementary information was obtained from the LCRC case dockets and the LCRC database for all cases included in the study. The LCRC dockets provided case particulars such as date of crime and composite quality indicators such as estimated resemblance between composite and suspect as judged by the witness. All categories of information acquired from the LCRC docket and LCRC database can be found in Appendix P-06. Table 5.6. lists all five forms of data collection and their respective return rate. For all four LCRC-related forms, 100%, or nearly 100%, of all cases could be collected. Only 54% of the forms sent to the investigating officers were returned.

Table 5.6. Numbers of Forms Included

	No forms collected	%
LCRC legal database[a]	576	99.7
LCRC docket[b]	578	100.0
LCRC fingerprint database[a]	578	100.0
IO form[c]	311	53.8
LCRC form[c]	578	100.0

[a] LCRC legal database and LCRC fingerprint database were both forms related to the electronic database available at the LCRC photographs.
[b] LCRC docket refers to the form in which data was collected from physical dockets available at the LCRC
[c] LCRC form and IO form refers to the questionnaire being filled out during the investigation by the police officer on duty

Data evaluation in archival studies is limited to information that can be found in police dockets, databases, or is collected by means of interview or questionnaire from police officers. While the former utilizes data reported independently of any study, the latter is not, raising the question of interpersonal expectancy effects (Wells, 2008). Because police officers know that the purpose of a study is in some regard to evaluate their performance, they might be tempted not be completely honest. Furthermore, police officers might be less than willing to participate and fill out additional paper work, and might consequently tend towards reporting as little as possible. Both tendencies would bias the resulting description, making it less objective. In an effort to try to measure these biases, several variables were collected twice. For example, investigating officers were asked to report other evidence available, such as fingerprints, that links the suspect to the crime. Fingerprint evidence can only link a suspect to a crime if his fingerprints were taken. Any fingerprint being taken is compared to a database of fingerprints, containing all fingerprints of convicts as well as fingerprints found at the scene of a crime. Whenever such a comparison is initiated, the LCRC database records the fact that a suspect's fingerprints were taken in this specific case. Therefore, one of the forms employed in the current study collected information on fingerprints found at a scene and suspects in the case whose fingerprints were taken. There should therefore be a sufficient amount of agreement between the IO form and supplementary information from the LCRC database. This comparison in my opinion is an interesting way to measure reliability of data collected in the field.

Procedure

The study was introduced to police management at the Western Cape Police Commanders Meeting on October 26th, 2005. The commanders were asked to authorize and support this study by encouraging station commanders and investigating officers to take part in it and to return the completed forms. The data collection phase was started on October 27th. This approach was chosen to ensure that all commanders were informed about this study, and therefore a sufficient amount of participation could be generated. Close to the onset of the study, station commanders of all stations involved in the study were contacted and asked to promote the study with their investigating officers.

Police officers at the face composite unit at the LCRC Cape Town filled out the LCRC form during and/or after constructing the composite with a witness. An IO form attached to a composite with a letter requesting its completion and return to either the LCRC or the university, was sent to the investigating officer who originally requested the composite.

Approximately four months after the meeting all investigating officers who had failed to return the forms were asked to return the outstanding forms. A research assistant phoned them, asking if they could assist in filling out the form, and emphasizing the importance of their continuous contribution. Eight months into the study, lists of all IO forms overdue were sent to the Provincial Detective Monitoring Division of the Western Cape and distributed to all relevant police stations. Such requests to return all outstanding IO forms were sent on a alternate month basis.

Parallel to the above, data were collected from the LCRC docket of each case and of the electronic database. The current legal status of each case was monitored, and once the case was closed the IO form was requested. Data capturing utilizing the police database was done by a police officer, while LCRC docket data were collected by a research assistant. Data collection was ceased about 12 months after the last composite was constructed, 24 months after data collection was started.

All data were entered into a data capturing mask. 10-15% or the data entered were controlled by a different person for input or coding errors, and a sufficient level of accuracy was found[8].

Results

Results in this section will be reported in relation to one another and to crime statistics of the same time period (National commissioner of the South African Police Service, 2006). Before doing so, I would like to highlight the implication of the chosen research method. The number of cases included in this study approximates 1/3 of the cases where a composite was constructed at the LCRC Cape Town during the study period. Naturally, when a witness is present most of the crimes included are contact crimes, described in the Annual Report of the South African Police for 2005/2006 as crime against a person. Also, with exceptions, composites are commonly constructed when the suspect is not known to the witness, that is, if the witness cannot identify the suspect as someone familiar, in which case the crime encounter is the only incident from which information leading to an arrest can be extracted. This leads

[8]There was a 97% rate of agreement between the two raters, with the error rate mainly stemming from initial glitches with the coding.

to a non-representative of composite cases with regards to the national level of different categories of crimes, a notion that will be discussed later in this chapter. Most of the data reported here was collected with the help of forms and lists. Due to the occasional misspelling of a case number, typing mistakes in the data sheet or other incidents, the attentive reader will find the mentioning of missing data in the following analysis, or that variables assessed with the same form or list have different total sample sizes. I deliberately abstained from compensation for them, that is by either excluding these cases from the analysis or replacing them with an average since the level of missing data in this study is very low. With respect to data from IO forms, the adjusted sample size of all 311 returned forms serves as a basis for relative percentages. Results of the data analysis will first be reported on a descriptive level before being related to relevant research in the subsequent discussion.

Case Particulars

Upon booking a case with the LCRC Identikit unit, investigating officers reported that in 83% of the cases only one witness present during the offence could provide information on the suspect(s). Only one witness was present during composite construction in 98% of cases, reducing the number of multiple witnesses willing or able to provide evidence in the form of a composite by 15%. 54% of the witnesses were female, ranging from 5 to 83 years of age and an average age of 32 years. Witnesses had a race distribution as described in Table 5.7.

Table 5.7. Race of Witnesses Constructing Composites

Race	N	%
Black	130	22.2
Coloured	255	43.5
White	197	33.6
Asian / Indian	4	0.7
Total	587	100.0

The race distribution of witnesses differs from the distribution of population groupings in the Western Cape as reported by the Municipal Demarcation Board (2001), $\chi^2_1 = 270.99$, $p < .001$. White witnesses were present 15% more often than suggested by an even distribution of crime among all race groups, indicating an increased prevalence of this group in being exposed to contact crime and/or reporting it.

Investigating officers reported to the LCRC that in 46% of the cases more than one suspect was reported to have committed the crime, with up to seven suspects having being present during an offence (LCRC docket). According to the LCRC form, 39% of cases contained more than one suspect. To clarify this issue, cases in which both forms were filled out were compared with respect to the number of suspects reported. Out of 490 cases in which a comparison was possible, consensus in the exact number of suspects was reached in 74% of cases. In only five cases witnesses reported more suspects while constructing the composite than the investigating officer had reported while requesting the composite(s).

In total 815 composites were constructed for which race, gender and age of the suspect on which the composite was based was estimated by the witness. Witnesses constructed more than one composite in 31% of all cases included, estimating the 'likeness' of the composite to the suspect's face to be around 75%. Police operators spent on average 46 minutes constructing the composite(s) with the witness(es). Witnesses reported that suspects on which the composites were based were 97% male, with an estimated average age of 28 years. They also reported a race distribution in those suspects depicted in the composites, as in Table 5.8.

Table 5.8. Race of Suspect on Whom Composites were Based as Reported by the Witness

Race	N	%
Black	480	58.9
Coloured	307	37.7
White	24	2.9
Asian / Indian	4	0.5
Total	815	100.0

Comparing the race distribution again to the census as stated above, Black suspects as described by the witnesses are overrepresented by 30% ($\chi^2_1 = 231.84$, $p < .001$), with both Coloured and White suspects being equally underrepresented with regards to regional population standards. On comparing witness and suspect race in each case it is found that in 58% of the cases a witness had to remember a suspect of a different race than her own.

Table 5.9 depicts the prevalence of different categories of offences within the sample as reported in the LCRC form, clustered by broader offence category. Robbery for example includes offences such as house robbery and hijacking/carjacking. Rape includes one case of sodomy. Theft includes house breaking. For 574 cases in which LCRC docket information on the crime type was available, comparison between those two forms of data gathering revealed a correspondence of 95% between reported crime types. The total numbers and percentages of crimes are displayed in relation to the national crime statistic of the study period as reported by the Annual Report for the South African Police Service of the year 2005/2006.

Table 5.9. Types of Crime (clustered)

	n	%	Incidence rate per 100000 of the population
Armed robbery	255	44	267.1
Robbery	102	17.7	140.1
Rape	75	13.0	111.0
Theft	73	12.6	1380.2
Murder	25	4.3	40.5
Assault	19	3.3	923.1
Attempted Murder	9	1.6	42,5
Fraud	5	0.9	
Kidnapping	5	0.9	
Attempted rape	4	0.7	
Attempted hijacking	1	0.2	
Attempted robbery	1	0.2	
Total	578	100.0	

As can be seen in the table, most crimes in which a composite was constructed were so-called contact crimes, a rate much higher than the 33.3% of recorded serious crimes as stated in the Annual Report. Armed robbery and robbery are the contact crimes with the highest incident rate within this study, in contrast to the national crime statistic in which assault is the contact crime with the highest incident rate. The Annual Report states that the Government's intention is to reduce specifically contact crimes, and that in this respect detection rate and court submissions increased over the last year. The most prevalent offence in composite construction, armed robbery, currently has a court referral rate of 10%. This means that in one out of ten cases in which a complainant reports an offence to the police, the investigation led to enough evidence for the prosecution to admit it to court. With respect to the offence with the second-highest incident rate in cases in which composites were constructed, common robbery, approximately one out of four cases was referred to court. Rape had a much higher referral rate than the latter two, with 42% of all reported cases being admitted to court. Court referral rates in this study will be discussed later in this chapter, relating these rates to the above mentioned numbers.

Witnesses stated that in two out of three cases the suspect(s) was/were in possession of a weapon during the offence. 84% of the witnesses interviewed were victim to the crime, that is, were either robbed at gunpoint, raped or otherwise directly affected by the action of the suspect(s). As an example a witness to a robbery might have stood very close to the robber with a knife held to her throat, asked to surrender her valuables. Witnesses estimated the typical duration of exposure to the suspect being five minutes.

With respect to other evidence available to the investigating officer, two separate measures were taken. As discussed earlier, the internal LCRC database provided information on fingerprint evidence for each included case. According to the LCRC databse, it was found that in 42.1% of the cases in which fingerprint evidence was sought, a positive fingerprint was found on the scene. A positive fingerprint is a fingerprint that can be used to incriminate or relieve a possible suspect. Of those 45 cases, 24 (53%) had no corresponding IO form information. 36% of investigating officers reported that no fingerprint evidence was found to link a suspect to a crime, and only 11% reported fingerprint evidence. Thus, in 24% of all cases in which fingerprint evidence was successfully retrieved from the scene and in which an IO form was returned, fingerprint evidence was reported to be helpful in apprehending a suspect.

Investigating officers were asked to provide further information on all available evidence in a particular case linking a possible suspect to the crime. For example, only body samples that were not related to the shop clerk of the shop where the armed robbery took place, but were related to the perpetrator were considered other incriminating evidence. Table 5.10 displays the distribution of such evidence based on the 311 forms returned. Information on how each category of evidence is defined can be found in the original form provided as Appendix P-05.

Table 5.10. Other Evidence Available in Included Cases

	N	%[a]
Fingerprint evidence	35	11.3
Property found in the possession of suspect	6	1.9
Video evidence	7	2.3
Body sample	7	2.3
Suspect confession	0	0.0
Evidence of co-accused	1	0.3
Identification of suspect	8	2.6
Other substantial evidence	3	1.0
Total of cases with other evidence	51	16.4

[a] Percentage here is based on the number of IO forms returned, which were 311 forms.

Table 5.10 shows that in cases in which a composite was constructed, fingerprints found either at the crime scene or on stolen property was the most prevalent other evidence in the investigation. Nevertheless, in 83% of all cases, the witness's memory of both incident and suspect is the sole evidence on which a police investigation can be based. Moreover, fingerprints can only be utilized if a match exists in the LCRC database, or if a person has already been identified as a suspect in the crime.

Time Delay

For each case, three dates were extracted from the LCRC dockets: date of crime, date of booking of the case with the LCRC by the investigating officer, and the date on which the the composite was constructed. Typically, investigating officers booked the case with the LCRC Identikit-Unit three days after the crime occurred (median), two days later than requested by policy as stated earlier. Table 5.11 shows the time interval between crime and composite construction.

The average delay was 27 days. Since some cases were booked 5 1/2 years after the crime took place, the median better reflects the delay as middle value of the distribution. A composite was typically constructed 13 days after the crime took place.

Table 5.11. Time Delay Between Crime and Composite Construction

N valid	577
Mean (days)	26.55
Median (days)	13.0
Std. Deviation (days)	91.35
Minimum (days)	0
Maximum (days)	1957

Distribution of Face Composites

Investigating officers were asked to specify how the composite(s) they requested was/were used in the investigation. Table 5.12 summarizes how the composites were distributed, and if it elicited any direct response to this distribution. A direct response here can be defined as a witness who saw the suspect running from the scene of the crime, or a person who has a neighbour Thabo looking just like the composite suggests and who came home late on the day of the robbery.

Table 5.12. Distribution of Face Composites

	N	%[a]	Responses to distribution
Published in newspaper	93	29.9	2
Shown on TV	22	7.1	1
Shown to informers or other specific people	198	63.7	7
Distributed to other police stations	194	62.4	0
Distributed in community where crime took place	165	53.1	0
Total of composites that were distributed	277	89.1	10

[a] Percentage here is based on the number of IO forms returned, which were 311 forms.

Composites were never distributed in isolation, but always with other information such as the date and location of the crime, the modus operandi or a description of the suspect; most often they were distributed with all three types of information. As can be seen from the table, composites were utilized in the investigation in nearly 90% of the cases. Out of 22 cases in which this distribution yielded a direct response, 10 (4%) responses were reported to be productive new leads.

Case Outcome

The investigative progress was measured 12-24 months after the incident, depending on when the case was included in the study. Table 5.13 shows legal status as reported in the LCRC database at the time of the conclusion of this study in October 2007.

In nearly 50% of all cases no conviction was reached either because no suspect could be apprehended or the case was withdrawn from court. It is difficult to compare these percentages to the performance rates in the Annual Report as stated above. The authors of the Annual Report state that no agreement could be reached regarding the measurement of conviction rate, and that the rate of 'detection' has to serve as a success indicator for police investigations. 'Detection' here means the percentage of cases referred to court, the number of cases withdrawn before court and the number of cases in which it was found that no crime was committed. Since 49% of all cases in this study are still under investigation or in court, no conclusive statement can be drawn regarding the performance outcome of cases involving composite construction.

Table 5.13. Legal Status of Composite Cases

	n	%
Under investigation	58	10.0
In court	224	38.8
Closed undetected	234	40.5
Closed withdrawn	41	7.1
Conviction reached	19	3.3
Missing	2	0.3
Total	578	100.0

As an alternative approach, an 'undetected rate' can be used as comparative measure. According to the Annual Report of the years 2006/2007 (National commissioner of the South African Police Service, 2007), 45% of all cases under investigation involving contact crime could neither be referred to court nor regarded as not in need of further investigation. For the years 2005/2006 a rate of 40% was reported. A similar rate was reported for cases in this study, with 40.5% being closed undetected, suggesting that cases involving composites adhere to the same tendencies as described for all contact crimes in South Africa.

5.4 Discussion of Study 4 and General Discussion

In 2004, when I started my collaborative research with the South African Police Service, I asked a member of the face composite unit how he estimates whether the composite he constructs together with the witness is a good-quality or a bad-quality composite. The answer I got was fairly disappointing: the only means he had estimating composite quality was by asking the witness's opinion. In this study this estimated 'likeness' yielded an average of 75%. Nevertheless, this does not mean that composites are 75% correct portrayals of the suspect face. It is a highly subjective estimate of the witness when comparing composite to remembered face. It stays unclear what the witness would consider a 100%-likeness, and if every witness has the same standard of what this 100% looks like. Estimating composite-suspect-resemblance in this manner helps to assess witness's satisfaction with the composite or willingness to continue with the construction process, and is a tool to initiate further changes in the composite by asking what would have to be done to make it a closer likeness. It is rather doubtful if the measure of 'subjective resemblance' in police practice can provide an objective estimate of composite quality.

A better way to define and measure quality of face composites is by measuring the contribution a composite makes to a police investigation. To measure quality this contribution has to be calculable, which proves to be very difficult in practice. If one were to do so, he/she would need to utilize composites in isolation from any other information in the case, and to for example record responses from its publication in a newspaper. Since a visual likeness without any additional information of where this person could have been seen and why he or she is wanted is seldom published, it is not possible to calculate the contribution of merely the composite. The utility of composites in police investigations can only be measured as an approximation of its quality.

Another way to establish face composite quality is to measure its "recognizability" or "visual likeness", as is done in experimental studies. In real life such an experiment can be constructed with cases in which a composite was constructed and the perpetrator was sentenced. In these cases it is highly likely that the person in the composite is the person convicted, and composite quality can be measured as stated above. Unfortunately, no permission could be obtained to use photographs of convicted perpetrators for study purposes, and any such attempt had to be relinquished.

Nevertheless, one outcome of this study is a better knowledge of the situation under which composites are constructed and which subsequently influence composite quality. The four broad areas that have been examined are witness characteristics, suspects' characteristics, crime particulars and composite utility. When comparing witness and suspect race it was found that in 58% of the cases a witness has to remember at least one suspect of a different race to her own. This is of specific interest because the significance of cross-race identification is one of the well-established and robust effects in research. In brief, it has been found that faces of a highly familiar racial group are remembered better than faces of a less familiar racial group (Meissner & Brigham, 2001b; Meissner et al., 2005; Sporer, 2001). For example, being exposed to predominantly White faces leads to better discrimination of White faces than faces from another race, such as Black or Asian, and vice versa. Since eyewitness testimony is often crucial in a police investigation, it is important to know when witnesses to a crime are likely to be accurate in remembering a face and when they are likely to make mistakes. The cross-race effect is one factor which makes performance differences between witnesses, at least to some extent, predictable (Ellis, Davies, & McMurran, 1979; Schmidt & Frowd, 2006).

Most crimes in these studies which required the construction of composites, were potentially violent and threatening contact crimes. It was found that in the majority of cases a person becomes victim to a crime in which a weapon is used to threaten the witness in either handing over property or enduring abduction. Most witnesses constructing composites were therefore exposed to a situation in which they experienced serious physical and/or mental harm or threat thereof, resulting in an emotionally stressful situation in addition to potential loss of property. Relating this to composite construction, witnesses most likely had a good look at the suspect and his face, and might therefore have a better memory of the suspect's face than witnesses in other kinds of situations. Indeed, Tollestrup, Turtle, and Yuille (1994) argue that victims and some witnesses to a crime are involved in the crime in a very different personal and emotional manner than participants in a laboratory setting. It might therefore be that affected witnesses and victims of a real crime are more accurate in their recollection of the event and the perpetrator. Tollestrup et al. for example found that robbery victims identified the suspect more often

(46.5%) than robbery witnesses (33.3%). Victims might process information about a crime differently to unaffected witnesses, leading to a potentially better memory of the suspect's face. Although the weapon focus effect in recognition memory is supported by a host of laboratory studies (Steblay, 1992), field studies on the effect of a weapon on subsequent identification did not find a detrimental effect (Behrman & Davey, 2001; Valentine et al., 2003).

Bennett (1986) stresses the opinion of many police officers that composites are more a marketing tool[9] than an actual investigative tool, therefore it is unlikely that they will be considered as the investigative tool whose sole presence will solve the case. Indeed this might be the case with some officers; however, cases in which composites are constructed show few alternatives in terms of investigative aids. The necessity to utilize composites in an investigation is demonstrated by the fact that only 17% of cases contained other evidence to link a suspect to the crime and to investigate the matter further.

Regarding the utilisation of composites by the investigating officers, results of this research were less than encouraging. The majority of composites were either distributed to other police stations, published in the community where the crime took place or shown to informers, with the latter yielding the most direct responses. This result draws a rather pessimistic picture of the usefulness of composites. Compared to the recognition rate of 20% of composites constructed in an experiment as suggested by Frowd et al. (2006), a response rate of 4% of published composites and related information that proved useful in either identifying a suspect or leading to other tangible evidence seems rather low. On the other hand, other evidence such as the much praised fingerprint evidence also did not fare as well as expected. Out of all cases in which fingerprints could be recovered at the scene of the crime or any related item and in which IO data was available, fingerprints helped furthering the investigation in 24% of cases included in this study.

Relating this research to the only other known published composite study in which surveys were sent to police officers (McQuiston-Surrett et al., 2006), a response rate of at least 50% in this study can be considered sufficient for a representative representation of police practice. South African crime reality might differ in some regard since its crime rate is higher than the international

[9]Bennett refers to some police officers' opinion that composites are being constructed to show the witness/victim that the investigation is proceeding. Therefore composites are used to satisfy the witness's need rather than to be published.

average, in general as well as regarding contact crimes like rape (Jewkes, Sikweyiya, Morrell, & Dukle, 2009). Crime statistics for the year 2004/2005 (National commissioner of the South African Police Service, 2005) showed a murder rate nineteen times higher than the murder rate of the United Kingdom, and nearly eight times higher than the international norm (Burger, 2007). On the other hand, South African police have a relatively open approach towards composite construction and use, only specifying in its policy (South African Police Service, 2003) that a facial likeness has to be established, leaving an open window for change and enhancement. In contrast to this, policies in the United Kingdom state that a composite has to be constructed within 48 hours after the crime has taken place, and that a cognitive interview has to be used (Association of Chief Police Officers, 2000, cited in Frowd et al., 2007a, p.564); the influence of both variables on composite quality has not yet been established.

Chapter 6

Evaluating the Evaluation – Experimental Comparison of Composite Quality Assessment Techniques

6.1 Introduction

Studies on face composites, as has been discussed in Chapter 3, use a variety of techniques to evaluate composite quality. These techniques differ in terms of cognitive processes required to complete them, relation to the face memory, experimental implementation as well as subsequent data level. In reviewing the studies to for example find out if one software system consistently outperforms other systems, it is necessary to compare the results of two studies on composite software programs. However, the one study might have measured composite quality with a line-up, the other study measured it with a likeness rating. Comparing results stemming from these two evaluation procedures would be to compare potentially incompatible results. Moreover, several studies utilize more than one evaluation technique, yielding different results with each technique. Studies described in Chapter 3 and 4 demonstrate this point: In several studies the independent variable showed a significant difference in one composite quality measurement, but not with a different measurement. This is not to say that using different measurements alone is responsible for differences in outcome. Different measurements might differ in both task difficulty and in drawing on different cognitive resources, both of which might have a large influence on the outcome. As discussed in Chapter 3, different cognitive aspects of the composite might be measured with different techniques. For example, likeness rating might draw on in-view feature-based comparison, compared to naming which draws on from-memory recognition of a well-encoded face instead. To judge if a composite is of good-quality is currently a rather subjective estimate. Estimating if one independent variable has a consistent influence on composite quality over several measurements and/or experiments is very difficult, if not currently impossible. A reason for this might be that there is no single, predominant theoretical nor is there a methodological definition of composite quality that is consensually employed by researchers. Composite quality is sometimes considered as having multiple dimensions (e.g., Frowd et al., 2006), and other times as being one-dimensional (McNamara, 2009). Yet, how these qualities relate to the method with which they are measured is often not all that clear. This chapter's experiment attempts to remedy the situation by comparing commonly used composite quality evaluation techniques.

In order to judge the influence of each evaluation technique on the estimate of 'composite quality', I consider it necessary to have comparable evaluation techniques. Only if one can find consistent trends for each evaluation technique can outcomes of studies on one specific aspect, utilizing different evaluation techniques, be compared. Currently a difference in results can be due to either the independent variable, the cognitive resource the evaluation technique draws upon, or task difficulty of the evaluation technique. In order to assess the influence of the evaluation technique on the result, the other two aspects should either be kept constant or be controlled. Using the same composites for each evaluation technique can achieve a constant influence of the independent variable on the outcome. Using the same foil photographs for evaluation techniques involving additional faces to choose from, might help adjusting task difficulty to a more comparable level. Introducing added measures for each task aids in controlling the influence of task difficulty on the outcome. The objective of this study is to compare the most common composite evaluation techniques by employing a research design adjusted to the above mentioned requirements.

The choice of composites to highlight quality differences between experimental conditions is crucial in this experiment. A non-significant outcome across all evaluation techniques would not be beneficial for the exploration of differences between the techniques. Using an old composite system that assembles facial parts manually while researching the influence of delay between encountering the target face and retrieving it on composites might not achieve a significant difference in composite quality, no matter which technique is used to assess this quality (Davies et al., 1978a, 1978b; Green & Geiselman, 1989; McNeil et al., 1987). In this case all evaluation techniques are likely to have the same result, that is, no significant difference between independent groups. If however, a composite system is used that can produce good-quality as well as bad-quality composites, researching the effect of an independent variable on composite quality is more likely to show an effect if it does exist. In this case different evaluation techniques are likely to show differences between those conditions, and might differ in the extent to which they show this difference, depending on aspects inherent in the evaluation technique. Therefore composites in this study have to be utilized that stem from a manipulation which results in different-quality composites, and as such hopefully consistent differences in evaluation techniques.

One manipulation which in former research has been shown to make a consistent difference is an in-view / from-memory comparison. As discussed in Chapter 3, it seems that the difference in memory conditions leads to a reliable difference in composite quality, which is necessary to study subsequent effects on quality assessment. As stated in Chapter 3, this manipulation has shown to have a consistent and large effect on the outcome, which in turn makes it highly suitable for this experiment. Nevertheless, this choice of manipulation might have an additional effect on the quality assessment. A face composite that was compiled in view of the target face can be compared and adjusted on a visual feature-by-feature basis, and this comparison might prove suitable for a later in-view composite evaluation. The participant-judges might compare each feature they find on the composite with the different target photographs in the matching task or with the single photograph in the likeness rating task. Because the evaluation task is an in-view comparison, the composites produced in-view of the original photograph might score higher in these tasks. On the other hand, this difference might be an informative result in and of itself, relating measurement tasks to cognitive concepts discussed in Chapter 2. As such I consider the use of composites made in-view and from-memory as stimulus material for this study suitable to the above stated objective, since differences in evaluation techniques will most likely be enhanced by this choice.

6.2 Methodology

Summary of Experiment

In the composite-construction stage one factor was manipulated, which was the memory condition under which composites were constructed. In one condition the participant-witness was shown a photograph of the target face prior to constructing a composite, thus had to construct the composite from memory. Since this study focuses on composite quality evaluation, a short delay in the from-memory condition was considered sufficient. In the other condition the operator constructed the composite with the photograph of the target face in view. These composites were constructed several days after the from-memory composites. In the following chapter these conditions are referred to as in-view and from-memory. The same Identikit expert constructed all composites. All composites were constructed from scratch and without time constraints.

In the composite-evaluation stage several different evaluation techniques were employed, all based on the same composites, and drawing from the same pool of target and foil photographs. Each different task was studied with a different, randomly assigned set of participant-judges. As common recognition tasks line-up identification, mugshot identification and reduction, matching, and naming were employed, in addition to which a likeness rating was used. The methodology of each evaluation technique will be described separately below. All evaluation tasks were introduced and discussed in Chapter 4.

A collaborative study design was required to implement the naming task. To mirror real-life conditions, targets had to be unknown to the participant-witnesses. Experiments not using the naming task usually use unknown target faces to avoid a familiarity with the target face for both participant-witnesses and participant-judges. However, in the naming task participant-judges have to be familiar with the target person in order to be able to recognize him from the composite. To ensure that target faces were unknown for participant-witnesses as well as most participant-judges, but known for naming participants, a collaborative study design was used. Steve Ross, then based at the University of El Paso, Texas[10], contributed to this study by choosing an initial set of faces of local celebrities for composite construction, as well as participating

[10] Steve Ross is currently Visiting Assistant Professor at Florida International University.

in the final choice of faces and by implementing the naming technique at the University of Texas at El Paso.

Composite Construction

Participants

A total of eight female participants for the composite production stage were recruited from the university campus by advertisement, with an average age of 21 years (19-23 years). All study participants either did this in partial fulfilment of their course requirement (choice of study voluntary) or were paid for their participation (ZAR 20, 2.60 US$). Participants were informed about the purpose of the study, and gave their informed consent. Those participants who took part in the reconstruction stage were not allowed to take part in the evaluation stage. Furthermore, those who participated in the pilot studies for the evaluation phase were not allowed to participate in the evaluation phase itself or serve as participant-witnesses.

Materials

Eight White male American celebrities were used as targets, with an age range of 18-55, having achieved fame either in sports or on TV. For each target several photographs with different views were collected, extracted from their background and subsequently either shown on a screen or as colour printout. Local American celebrities were chosen to make it unlikely for participant-witnesses as well as participant-judges in South Africa to be familiar with them beforehand. Stimulus as well as foil photographs can be found in Appendix E-09.

Sixteen White male targets were chosen from a list provided by the US collaborator, all of which were not well-known in South Africa. The celebrities were chosen to be nationally rather than internationally known, as are football players, race car drivers and winners of recent national song contests. A pilot-study was conducted to inform that selection. Each target face of the preliminary set was printed on a colour 10x15cm cardboard, and 20 participants not involved in the rest of the study, indicated which ones were known to them. For the same

16 target faces recognition was tested with 23 participants of the University of El Paso, Texas. From this initial choice of 16, the targets which were least known in Cape Town and best known in El Paso were chosen as study stimuli. To further ensure that participant-witnesses did not know the target prior to composite construction, they were asked if they knew the target, and subsequently given another target as stimuli if they did. Likewise, participant-judges in El Paso were given the target photographs after completing the uncued naming task to test for target familiarity.

Because celebrity photographs on the internet seldom come in full-frontal view with a neutral facial expression, an individual photograph might not display all facial characteristics needed to construct a composite. If for example the target is always portrayed smiling, a witness-participant might be unable to judge the shape of his lips if shown with a neutral expression. Likewise, a ¾-view target picture might not reveal that the target has protruding ears. For that reason two photographs of each target were used for the composite constructing phase, at least one of which had to satisfy the following criteria: approximately frontal view, closed mouth, no accessories (like hats) and no disguised view (e.g., a microphone in front of the mouth). One of the inclusion criteria for targets was the availability of a photograph as described above, in colour and with a size of at least 300x300 pixels.

For the mugshot tasks, the likeness rating task and the line-up task, six foils for each target were selected. Adhering to the match-to-description selection strategy (Luus & Wells, 1991), the most commonly described facial characteristics reported by the participant-witnesses and the composite system operator (henceforth operator) were used and guided the experimenter in the choice of foils. Foils were chosen either from a set of unstandardised photographs of unknown men from a large database of faces, or from the internet, choosing similar American as well as European celebrities relatively unknown in South Africa. Celebrity faces were taken from the same populations than the target celebrities: sports and TV personalities. For example, for the target Troy Aikman, John Elway and Joe Montana were chosen as foils, all three of whom are American football players. Unstandardised foil photographs were used in order to match target photographs. In the evaluation tasks approximately frontal views of the target and of six foils were used for each target face, for all of which only head and neck were visible. Furthermore, all photographs were post-processed with Adobe Photoshop to attain an equitable level of diversity and similarity between photographs., and standardised to 300x300 pixels.

Identikit 2000 was used as the composite construction system to increase the comparability to other studies within this thesis (Chapter 4) and police practice in South Africa (Chapter 5).

Procedure

Participants were tested individually throughout and randomly assigned to the target person. Upon their arrival each participant was asked to sit in front of a computer screen and follow the instructions that were being presented to her. She was then asked to remember the person she was about to see on the next screen for a later task, after which the target photograph was shown for 20 seconds. After the face disappeared, participants were asked to give a brief description of the target, which was subsequently utilized for composite construction as well as for choosing foils for each target as stated above. As visual filler task, a digit span test was administered for five minutes, as described in Chapter 4.

Participants were then informed that they had to construct a composite of the person they had just seen. The operator chose the starting face according to the description provided by the participant earlier, and demonstrated the tools and choices of Identikit before starting with the composite construction. The operator was blind to the identity and appearance of the target, and participants were informed of this and told that all information had to come from their memory. Participants were told that they could take as much time as they needed, and make as many changes as they felt necessary until they were satisfied with their construction. Similarly to the studies described in Chapter 4, Adobe Photoshop was used as additional graphic's program to make changes to the composite whenever Identikit did not provide the necessary tools.

For the in-view condition the same procedure was applied, only that the operator filled the role of the participant-witness, and constructed a composite of each target after the from-memory composite was constructed. Participants took an average of 35 minutes to construct the composite, whereas the operator took an average of 40 minutes in the in-view condition. A total of 16 composites were constructed of eight target faces, half in-view and half from memory with the help of a participant-witness. Composites and related target photographs can be viewed in Appendix E-08.

Composite Evaluation – Likeness Rating Task

Participants

37 students of the University of Cape Town were recruited. 86% of participants were female, 68% White, and participants were on average 22 years old (range 19-49 years).

Materials

Photographs for both targets and foils as stated above, as well as composite images were used.

Procedure

To make the presentation design more accessible, each evaluation task is presented here in a figure containing a schematic exemplar on how composites and photographs were presented. Figure 6.1 shows a pair of composite-target images presented on the computer screen for the likeness rating task.

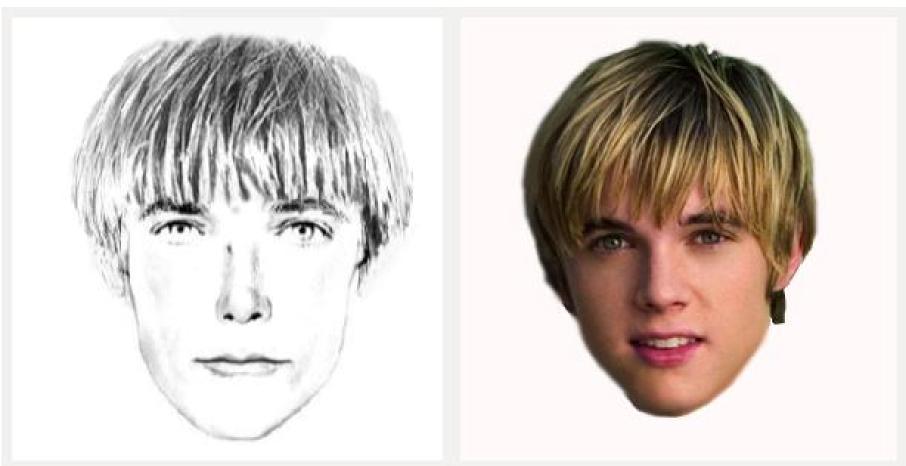

Figure 6.1. Example of a Composite-Photograph Pairing in the Likeness Rating Task

The procedure was similar to the likeness rating task in Chapter 4, requiring each participant to give a likeness rating for each of the photograph-composite combinations. In addition to rating composites in terms of their likeness to the target picture, composite likeness was also rated in relation to the six foil pictures related to the target. A 101-point-scale was used for each likeness rating.

Composite Evaluation – Line-Up Task

The procedure employed in this task was similar to that of the line-up task in Chapter 4. As mentioned earlier, two composites were constructed for each of the eight target faces, one in-view and one from-memory. Also, two line-up conditions were utilized, target-present and target-absent. Each participant saw only one line-up per target in order to avoid recognition of a previously seen line-up and subsequent influence of prior decisions for this line-up.

Participants

138 students of the University of Cape Town were recruited. Participants were on average 21 years old (range 19-49), 65% White and 78% female. Every participant saw only one line-up per target, so eight line-ups in total. The choice of target, memory condition, line-up condition and sequence of presentation was randomised.

Materials

Six-person line-ups were constructed, target-present as well as target-absent. For the target-absent line-up, the foil most similar to the target by judgement of the experimenter was chosen as target-replacement. All line-ups were pre-tested for line-up fairness (see Chapter 4 for a description of the procedure). An effective size of at least 3.2 (ø 4.2) as well as chance-level choosing rate for the target face in the target-present line-up was considered sufficient to proceed with the evaluation.

Procedure

Participants came to the laboratory and were asked to open an online link to access the instructions and line-up presentation. The instructions specified that the composite they were about to see was constructed from a target face that might or might not be in the line-up to follow. They were told that they could choose one of the displayed faces if they recognized the person as the person depicted in the composite, but they could also reject the line-up if they thought the person depicted in the composite was not present. Following the instructions, the composite was displayed on the screen for 10 seconds, after which the line-up was shown. Figure 6.2 shows an illustration of this task.

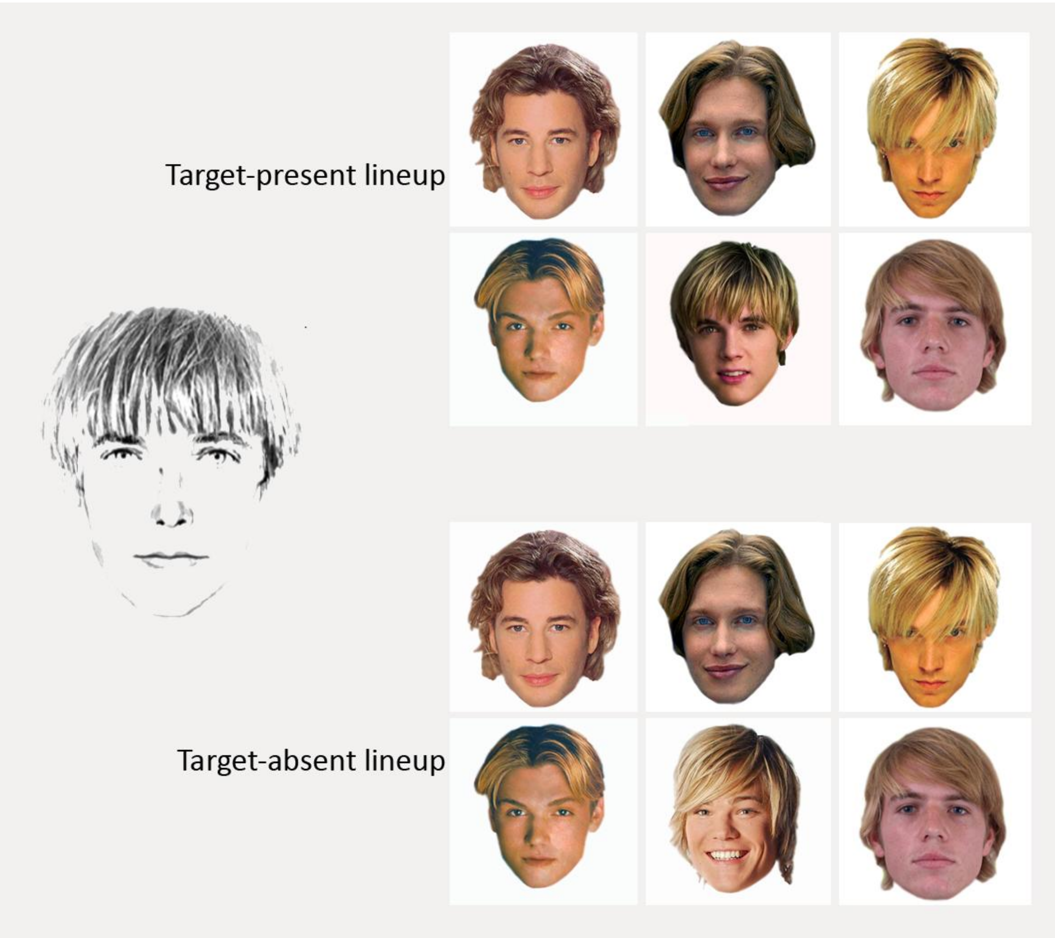

Figure 6.2. Example of a Composite-Line-Up Pairing in the Line-Up Task

All line-up members were presented simultaneously to keep the diagnostic value similar to other identification techniques like matching and mugshot, as well as to avoid a floor level identification rate. The stimuli were presented on the computer, the program accessed via the internet and subsequent data saved electronically.

Composite Evaluation – Matching Task

Participants

34 students of the University of Cape Town were recruited. The average age of these participants was 21 years (range 19-39 years), 68% were White and 76% female.

Materials

Composites and target photographs were printed with a high-resolution colour laser printer, and each mounted on a 4x6 cm cardboard. Target photographs were assigned numbers 1-6 in the bottom right corner, whereas composite cards contained the letters a-p. The answer sheet contained instructions and response options, and can be found in Appendix P-07.

Procedure

Participants were given an answer sheet and an envelope containing the above mentioned image cards. The first page of the answer sheet contained instructions for this task. The participants were asked to first spread out the photographs in front of them, and to have a close look at each of them. Figure 6.3 displays the array of target photographs, with one of the composites to match.

Thereafter the participants had to inspect the pile of composites, look at the first one only, and decide which of the faces in the photographs was the closest match to this composite. They were asked to go through the pile of composites in a serial manner, each time making a choice and putting the composite next to the photograph they chose. They were told that there could be none, one or multiple composites assigned to each photograph. After they had chosen a photograph for each composite they were asked to report this choice in the

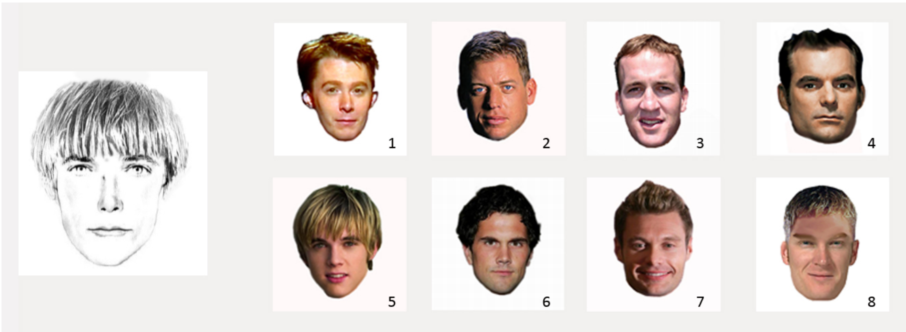

Figure 6.3. Example of a Composite and Target Photographs in the Matching Task

answer sheet. After all 16 composites were assigned to one of the photographs the participants were thanked and dismissed.

Composite Evaluation – Mugshot Tasks

In this subsection both mugshot identification and mugshot reduction will be described. Mugshot identification was discussed in Chapter 4, and is one of the identification tasks commonly used in experiments measuring composite quality. Mugshot reduction however is a novel task. The implementation of this task is a logical consequence of the discussion on the utility of composites. Composites cannot only be used to focus an investigation on one or more specific suspect(s), but also to exclude other suspects from further investigation. Following this thought, it is just as important to reject the composite as being a specific person as it is to identify a person from a composite. This thought is already encroached on the line-up task in which target-absent line-ups are presented and ought to be rejected if the person depicted in the composite is not present. Therefore two mugshot tasks were implemented in this study. Mugshot identification hinges on the identification of multiple targets possibly displayed by the composite. Mugshot reduction hinges on the exclusion of targets from the set of targets possibly displayed by the composite.

Participants

For the mugshot identification task, 36 students of the University of Cape Town were recruited. Participants were on average 21 years old (range 19-23), 56% White, and 78% female. 36 participants taking part in the mugshot reduction task were on average 21 years old (range 19-33 years), 65% White, and 92% female.

Materials

The same target and foil photographs were used as in the line-up task, each being displayed with a visible number in the left bottom corner of the photograph. Since a total of 56 photographs cannot be presented on a 17" computer screen without loss in photographic quality they were projected onto a screen. The screen was 1m wide and 1.40m high. Participants were seated in a close half-circle around the screen, with the middle chair directly opposite the screen being 2.4m away from the screen. Photographs were presented in seven rows, each containing eight images. On the screen, photographs were 12cm wide and 20cm high. The placement of targets and foils was randomised for each group. Figure 6.4 shows an illustration of this task.

Figure 6.4. Mugshot Screen and Accompanying Composite in Both Mugshot Tasks

An A5-booklet was compiled for each participant, containing all 16 composites in randomised order. Each page contained one composite. Instructions on the cover page and on each composite page differed, as can be seen in Appendix P-08. Mugshot identification participants were instructed to choose who of the persons displayed on the screen they thought the composite depicted. They were then asked to make a 1st, 2nd and 3rd choice. Each composite page had a heading repeating this instruction in a simple question: "Which photographs would you match to it?" Participants in the mugshot reduction task were given a table containing rows and columns of numbers (matching to the photographs on the screen), and asked to "cross out the numbers of the photographs you think do NOT depict the man in the composite below". They were told they could exclude/include as many photographs as they liked. The answer sheet for the mugshot reduction task can be found in Appendix P-09.

Procedure

Participants were invited to the laboratory in groups of six, and were led to a semi-dark room with a screen, chairs surrounding it, with just enough lighting to see both the composites in the booklet and the photographs projected onto the screen. They were given one of the booklets described above with a cover page containing instructions and all subsequent pages containing one composite each. The experimenter emphasized that each composite was constructed from one of the faces they saw on the screen in front of them, and they had to judge which of the persons in the photographs might be depicted in the composite. They were told that they should work in a sequential manner, that is, not skip one page of the booklet or scroll through the booklet. They were told that it was possible that one or more of the composites related to or portrayed the same person, so they could choose or cross out the same photographs several times. They were then asked to start, with the screen constantly showing the same image as depicted in Figure 6.4. At the end of the task, they were thanked and dismissed. There were no time constraints to this task.

Composite Evaluation – Naming Task

Participants

35 students of the University of El Paso, Texas, were recruited as participants. Participants were on average 20 years (range 18-31 years) of age , were 54% male and 86% Hispanic.

Materials

A presentation was used to display each composite as well as the instructions. The presentation included firstly a set of instructions, followed by a composite on every page. The answer sheet (Appendix P-10) contained instructions, demographic data as well as response options.

Procedure

Participants were informed that the composites they were about to see were reconstructions from members of the public in the US, be it actors, sport stars, singers or entertainers. They were asked to guess who it might be. They were also informed that a celebrity could be (but didn't have to be) depicted several times and could therefore name them more than once where applicable. They were then shown all 16 composites sequentially. Participants had to report each answer on the answer sheet provided. As can be found in several naming tasks (see Chapter 4), participants could either provide a name or other identifying material. These answers were later coded into "named" or "not named" by the collaborator. The composite was considered "named" if the participant could supply the experimenter with unique identifying information, as for example the name, a specific film role or when this person won a certain award.

Participants were subsequently asked to identify the target photographs. If they could not identify at least five of the eight targets their data was excluded from further analysis. After identifying the target photographs, participants were again asked to identify the composites, a procedure named 'cued naming' by Frowd et al. (2007a).

6.3 Results

Results will first be presented separately for each technique, with both traditional and additional measurements. Results will subsequently be presented in comparison to one another in the discussion. A common inferential analysis was used for all tasks, in addition to individual analysis. Accounting for the partial dependency of data, the McNemar test as an exact test based on a binomial distribution was utilized. Since the McNemar test for small samples is done as an exact probability test rather than a test through the chi-square distribution, the likelihood-coefficient is reported instead. The Mantel-Haenszel approach as suggested by Kuritz, Landis and Koch (1988) was used for the logarithmic odds ratio estimate ($OR_{(ln)}$). $OR_{(ln)}$ are used in order to display a better comparability with the effect size estimate in the likeness rating.

Likeness Rating Task

Traditionally, likeness between the target photograph and the related composite(s) is rated along a scale, and a significance test is used to compare two or more experimental conditions. In this study, the in-view ratings were be compared to the from-memory ratings, as can be seen in Table 6.1.

Table 6.1. Likeness Ratings for Target-Composite and Foil-Composite Comparisons

	Target		Foil	
	Mean	SD	Mean	SD
From memory	28.78	27.93	24.26	17.98
In view	53.88	29.69	34.53	17.95

A paired sample t-test including only values from the target-composite comparison showed a significant advantage for the in-view condition, ($t_{(1,295)} = 13.47$, p < .001). Cohen's d for this comparison is d = .87. In-view composites in the foil-composite comparison were also rated significantly more like the presented foil then were from-memory composites, $t_{(1,295)} = 10.52$, p < .001, d = .57). In other words, both the target and the foil average were rated to have a better likeness towards the in-view composite than the from-memory composite. Since none of the foils are depicted in the composite, it is surprising that here too, likeness is rated better when the composite was constructed in view of the target. On the other hand, if it is assumed that composites display a type likeness rather than a specific person, the average foil, chosen to match a description of the target, might contain enough elements of the target to warrant this result. An alternative way of analysis, notably a more comparable analysis in relation to the other techniques of composite quality assessment, is to convert likeness ratings to a categorical data level. Using stimulus material already available for other techniques, I defined 'hit' for this technique as being a likeness rating of composite-to-target above a likeness rating of composite-to-foil-average. In other words, a participant would have to rate a composite more similar to the target than the average foil to code it as a 'hit'. A 'miss' is defined as a likeness rating of a composite-photograph combination that is larger for the average foil than for the target. In cases in which both likeness ratings were equal, the result was counted as a 'miss', as is done in the naming task if a participant cannot settle on one identification. A total number of 29 comparisons, that is, 4.9%, resulted in equal likeness ratings.

Table 6.2. Likeness Frequencies of Hits and Misses in Relation of Target and Foil Average

	Hit		Miss	
	n	%	n	%
From memory	136	45.95	160	54.05
In view	228	77.03	68	22.97

As can be seen in Table 6.2, in-view composites were rated more likely to be the target than the average foil more often than composites constructed from memory. This difference was significant ($\chi^2_{(1)} = 61.71$, $p < .001$, $OR_{(ln)} = 1.372$). This result is interesting because overall, likeness ratings were higher in the in-view condition, both for target-composite and foil-composite comparisons. However, the highest average rating was found in the in-view target-composite comparison, contributing to more hits in the in-view comparison than misses. Participants rated target-composite combinations higher 3.35 times more often than foil-composite combination. Taking both analyses together, in-view composites seem on the one hand to spur higher ratings in general. On the other hand, they seem to affect more the target-composite rating than the foil-composite rating in this condition.

Line-Up Task

The line-up task provides several traditional measurements, all of which are shown in Table 6.3.

In-view composites led to more targets having been chosen in the target-present condition ($\chi^2_{(1)} = 14.34$, $p < .001$, $OR_{(ln)} = .76$), but also to more target-replacements being chosen in the target-absent condition ($\chi^2_{(1)} = 3.22$, $p <$

Table 6.3. Line-Up Identification Frequencies of Choices

	Target-present						Target-absent					
	target		foil		rejec-tion		replace-ment		foil		rejec-tion	
	n	%	n	%	n	%	n	%	n	%	n	%
From memory	51	17.47	12	41.1	12	41.1	17	6.25	12	44.1	13	49.6
In view	82	31.18	10	28.4	80	30.4	29	10.47	15	54.1	98	35.3

.001, $OR_{(ln)} = .56$). The comparison of diagnosticity (Wells & Lindsay, 1980) calculated as per Tredoux's recommendations (1998) showed no significant diagnosticity difference between from-memory and in-view composites ($\chi^2_{(1)} = .02$, p = .89). Looking at the diagnosticity itself, from-memory composites achieved a diagnosticity ratio of d = 2.79, in-view composites achieved a ratio of d = 3.00. One explanation is a possible shift in decision criterion (Ebbesen & Flowe, 2002). From-memory composites were more often rejected than in-view composites in the target-present line-up ($\chi^2_{(1)} = 7.31$, $p < .001$, $OR_{(ln)} = .48$) as well as in the target-absent line-up ($\chi^2_{(1)} = 11.45$, p = .015, $OR_{(ln)} = .59$). Summarizing these results, in-view composites seem to increase the number of correct choices in the target-present line-up by a general increase in overall choices, both in the target-present and target-absent line-up. Calculating the correct choices overall, in-view composites did not show an advantage over from-memory composites.

Matching Task

As discussed in Chapter 3, many prior studies have used analysis of variance as parametric statistic for interval data to compare data from different experimental conditions by adding hit/miss over all targets of one condition. I abstained from this approach, and kept the data dichotomous. As can be seen in Table 6.4, matching data shows similar relations between in-view and from-memory composites to likeness ratings.

Inferential statistics showed a significant difference between in-view and from-memory composites, with in-view composites yielding more than double as many hits as from-memory composites ($\chi^2_{(1)} = 51.43$, p = .005, $OR_{(ln)} = 1.28$).

Table 6.4. Matching Frequencies of Hits and Misses

	Hit		Miss	
	n	%	n	%
From memory	75	27.57	197	72.43
In view	157	57.72	115	42.28

Mugshot Identification Task

Like the matching task, the musgshot identification task has a dichotomous hit/miss outcome. Table 6.5 shows the frequencies with which participants identified the target as the person depicted in the composite in the first choice category or in one of the three possible choice categories. Please note that 56 photographed persons were available to choose from in the mugshot image.

Table 6.5. Matching Frequencies of Hits and Misses

	Only 1^{st} choice				All three choices			
	Hit		Miss		Hit		Miss	
	n	%	n	%	n	%	n	%
From memory	24	8.33	264	91.67	62	21.53	226	78.47
In view	63	21.88	255	78.12	117	40.63	171	59.37

As can be seen in the table, participants identified the target as first choice in the from-memory composites in 8% of all cases, whereas composites constructed in-view of the target were identified in 22% of all cases. This difference was significant ($\chi^2_{(1)} = 21.24$, $p < .001$, $OR_{(ln)} = 1.13$), as was the difference for all three choices ($\chi^2_{(1)} = 24.82$, $p < .001$, $OR_{(ln)} = .91$). Please note that data for these comparisons is related, that is, all three choices naturally include the first choice. To increase comparability with other measures, the 3-choice-data should be chosen since this data is more related to the notion of type likeness, and to the number of choices per alternative[11].

[11] In a line-up, participants have to choose one out of eight photographs. In a three-choice scenario mugshot task, participants have to choose 3 out of 56 photographs, so have an approximately 1 in 19 chance of picking the target by chance, instead of a 1 in 56 chance if only the first choice was considered.

Mugshot Reduction Task

Since mugshot identification is a novel technique, no traditional methods of analysis have to be abided by. For the mugshot reduction task, two related measurements were analysed: the number of 'hits', and the set reduction rate. 'Hit' in this case refers to the target photograph being included in the set of photographed faces being potentially depicted in the composite; 'miss' refers to the target being excluded. As can be seen in Table 6.6, target photographs were included in the set nearly twice as often when the decision was based on in-view composites than when it was based on from-memory composites.

Table 6.6. Mugshot Reduction Frequencies of Hit and Miss

	Hit		Miss	
	n	%	n	%
From memory	92	31.94	196	68.06
In view	167	57.99	121	42.01

The second measure of this task relates to the number of photographs included into the set. The usefulness of this measure is related to the inclusion or exclusion of the target photograph. The better the 'quality' of the composite, the smaller the included set should be, while containing the target photograph. Table 6.7 shows the number of photographs included in the set, subdivided by if the target photograph was included into the set or excluded. Please note that, different to prior tables, groups 'hit' and 'miss' are related to Table 6.6, therefore percentages are not related to the number of hits/misses, but to the number of photographs in the set.

Table 6.7. Mugshot Reduction - Number of Photographs Included in the Set

	Number of photographs included if...			
	Hit		Miss[a]	
	n	%	n	%[a]
From memory	9.33	16.67	27.50	49.11
In view	13.41	23.95	20.35	36.34

[a] Percentage here is the number of included photographs relative to the complete set of 56 photographs.

As can be seen in the table, when deciding which photographed faces could be depicted in the in-view composites, participants included about 13 photographs when the target photograph was included, while from-memory composites resulted in the inclusion of only 9 photographs. When the target was excluded, that is, when participants decided the target photograph could not depict the person in the composite, about 20 and 28 photographs respectively were included. Relating this to police terminology, witnesses seem to consider more photographs be possible depictions of the composite if the target was rejected.

Combining both measures, a weighted identification measure was calculated. In order to do so, each hit and miss frequency was weighted by the reciprocal of the percentage of included photographs. This would for example mean that from-memory hits were multiplied by 83.33%. The resulting weighted identification rates showed a non-significant result ($\chi^2_{(1)}$ = 13.46, p = .084, $OR_{(ln)}$ = .76).

Naming Task

For the naming task, participants were asked to identify the person depicted in the composite twice. The first time they were asked to identify the person as someone in the American show business or sport industry (uncued naming). For the second (cued naming), they had been shown the original target photographs, which might have given them an idea whom these composites were depicting.

Table 6.8. Naming Frequencies of Hit and Miss

	Uncued naming				Cued naming			
	Hit		Miss		Hit		Miss	
	n	%	n	%	n	%	n	%
From memory	0	0.00	280	100.0	15	5.36	265	99.05
In view	1	0.01	279	99.99	50	17.86	230	82.14

As can be seen in Table 6.8, uncued naming yielded only one correct identification response. After showing the photographs of the target faces, however, 18% of in-view composites and 5% of from-memory composites were identified, $\chi^2_{(1)}$ = 3.43, p < .001, $OR_{(ln)}$ = 1.37. On average, eight alternative identities were suggested for the uncued from-memory composites, whereas on average six alternatives were falsely identified for the uncued in-view composites. As for the cued composites, on average six alternatives were suggested for both categories.

6.4 Comparison of Methods and Discussion

Focussing first on the likeness rating task, this study found a significant advantage of in-view composites when rated against target photographs. This result is in line with studies measuring composite quality with a likeness rating while investigating the difference between in-view composites and from-memory composites (Brace, Pike, & Kemp, 2000; Brace et al., 2006a; Bruce et al., 2002; Christie et al., 1981; Ellis, Davies, & Shepherd, 1978; Ellis et al., 1975; Laughery & Fowler, 1980). This finding supports the argument of all authors that composites constructed in-view of the target should be of better quality than composites constructed from the memory of the same face. Furthermore, in-view composites should certainly perform better in an evaluation task that involves in-view comparison of composite and target face, as is the case in the likeness rating task.

However, as an additional measure in the likeness rating task, the similarity between composite and foils was rated too, using the same methodology and the same participants as the composite-target rating employed. Here too, a significant increase in rated likeness of in-view composites compared to from-memory composites was found. In other words, the average foil was rated to be more similar to the in-view composite than to the from-memory composite. This result cannot be intuitively derived from the idea that the quality of a composite is defined by its accuracy of features and configurations or its overall recognizability as the target face.

This result however, gives rise to a new notion of what composite quality, as measured with currently used quality assessments tasks, entails. In a line-up task, recognition rates are viewed as having two separate aspects, one being inherent to the memory of the target face (diagnosticity) and the other one being inherent to the task itself (decision criteria). Extending this concept to composite quality measurement, these two seperate, independent aspects may be inherent to the composite. One aspect is the diagnosticity. The more diagnostic a composite is, the more it enables a random viewer to recognize the target face, and to reject other faces. The second aspect entails a criterion shift within a random viewer (Ebbesen & Flowe, 2002; Flowe & Ebbesen, 2007; Meissner et al., 2005), and is called decision criterion. The lower the criteria, the more likely a random viewer is to choose, irrespective if this is the target face or another face. Also, the lower the criteria, the more alike a random viewer will rate the composite to a photograph, irrespective if this photograph depicts the target or another person. I suggest that the decision criterion in composite research is not only related to the task at hand, but also to the composite itself. In other words, some composites will prompt a random viewer to judge more often than others. In this discussion I will relate both concepts to the study findings. Relating these concepts to other research and theoretical considerations is part of the conclusion of this thesis.

Coming back to the likeness rating, participants rated the in-view composite and the photograph more alike, irrespective if the photograph depicted the target or a foil. This measure does not answer if in-view composites contain a higher diagnosticity, but it suggests a lower decision criterion. Decision criterion in this case means that participants are more willing to judge composites a good likeness, and thus give higher ratings. Comparing likeness ratings of target-composite and foil-composite combinations and combining them into a dichotomous measure showed a significant difference in that target-composite combinations in the in-view condition were rated to have a higher similarity to the target face than the foil average. Relating this measure to the measure of diagnosticity in line-ups, participants here were more accurate by assigning a higher likeness rating to the target than the foil. If diagnosticity is an expression of accuracy not biased by the decision criterion and including target-absent choices, one could deduce that in-view composites showed a higher diagnosticity in the likeness rating.

The matching task as well as the mugshot identification task showed a significant advantage of the in-view composites over the from-memory composites, thereby replicating the findings of for example Davies et al. (2000). Inherent in each of those tasks is the necessity for the participant to choose, that is, to identify a photographed face as being depicted in the composite presented. A possible criterion shift, defined as willingness to choose, cannot be measured with a task in which the participant is forced to choose, and in which the target is always present.

The line-up task is a very well-defined method in eyewitness testimony, with measurements over and above the mere frequencies of hits and misses. Starting with the frequency of hits, from-memory composites in the target-present condition accounted for a hit rate of 17.5%, whereas in-view composites accounted for a hit rate of 31.2%. The target was chosen significantly more often when the decision was based on in-view composites than when it was based on from-memory composites. Comparing hit rates found in this study to hit rates expected by chance, however, draws a different picture as to the quality of composites. The calculation of hits by chance depends on the number of assumed alternatives. Using all available alternatives, a hit rate of 16.7% is predictable if hits are solely based on chance. If chance level is based on the averaged effective line-up size, measured in the pilot study, a hit rate of 27% is to be expected. Assuming that the effective size measure is a more reliable estimate of how many real choices are available in a line-up, both composite conditions did not achieve a hit rate much higher than chance level. However, this result might relate to line-up task difficulty more than to composite quality.

Focussing again on the differentiation between diagnosticity and decision criterion, both identification rates and rejection rates of target-present and target-absent line-ups have to be considered (Ebbesen & Flowe, 2002). Both target-present and target-absent line-ups showed a significantly higher rejection rate for composites constructed from memory than when constructed in view of the target. This result suggests that participants were more willing to choose when they were exposed to an in-view composite, supporting the interpretation of the likeness rating result that both diagnosticity and willingness to choose are an inherent quality in composites. In-view composites seem to trigger a lower criterion for making a choice, irrespective if this choice, identification or higher rating, is related to the target's identity.

Diagnosticity, as defined by Wells and Lindsay (1980) and refined by Tredoux (1998) is the overall accuracy of the witness's decision. In the context of line-ups in composite research, diagnosticity relates to the ability of a random person to recognize the target in a set of faces, and to correctly reject another set which does not contain the target, solely based on the information provided by the composite itself. Both in-view and from-memory composites have a logarithmic diagnosticity of above one, meaning that both categories enable the participant to more accurately choose or reject a line-up than a rate dictated by chance. However, no significant diagnosticity difference was found between in-view and from-memory composites. Combined with the criterion shift found, it can be hypothesized that the advantage found for composites constructed in view of the target related more to an increase in the willingness to make a choice (or give a better rating) than an increased ability to differentiate the target from a pre-selected foil.

Floor-level results in the spontaneous naming task indicate, as discussed in Chapter 3, a need to establish task difficulty before the experiment to avoid these results. The request to identify the target as a person of the general public resulted in a naming rate of close to zero, as has happened in other studies using a comparable instruction (e.g., Frowd et al., 2007a). The increase in correct identification of the target in the cued-naming task could be attributed to a criterion shift rather than an increased diagnosticity due to a more recognizable composite. Participant-judges could have been more willing to choose, and therefore by chance arriving at more correct identifications. Since there is no target-absent condition in the naming task, the influence of the willingness to identify a person on the naming rate can only be estimated. Experiments by Brace et al. (2000) as well as Davies et al. (2000) also found an effect of the in-view/from-memory manipulation on the frequency of correct identifications. Participants in these studies were asked to identify persons they know from the departmental teaching staff, reducing the persons the composite could depict to a limited pool size, as might be the case in the cued naming task. In addition to the naming of persons participants had to be familiar with, Davies et al. incorporated a novel target-absent condition into their naming task, asking participants to name composites of persons they could not be familiar with. In their study, more correct names than incorrect names were recorded over all conditions. Contrary to the study reported in this chapter, Davies et al. found the correct naming rate to be significantly higher in the in-view condition than in the from-memory condition, but no higher choosing rate (correct and incorrect identifications) over both target-present and target-absent line-up.

In summary, it can be deduced that in this study composite quality evaluations measured at least two separate components of composite quality: diagnosticity and decision criteria. For those measures that do not entail a decision criterion, such as matching, mugshot identification, or target-present line-ups with a forced-choice, diagnosticity is dependent only on task difficulty and properties inherent in the composite. Composite quality as measured by this and previous studies does not seem to be a straight-forward concept like 'physical accuracy' or 'recognizability', but entails more distinct components that all contribute to the performance outcome of the constructed composite. Therefore I refrained from calculating a single metric to compare the outcome of the different measurement tasks. Relying on this one number to describe composite quality would obscure the fact that there is no single attribute called composite quality. A further discussion will conclude this matter in Chapter 7, and elaborate on the relation of theoretical concepts and composite evaluation task methodology to outcome. One conclusion that can already be drawn from this study is that evaluation methods, even if they rely on the same material, are inherently different, leading to different outcome measures previously defined as ’composite quality’.

Chapter 7

Conclusion

In this thesis I have covered a broad range of the many, varying aspects of composite research. In this final chapter I would like to focus on several of the most important aspects of composite construction in terms of its relevance for future research and policy formulation.

First, I will review the arguments made in this thesis for featural and holistic processing during composite construction, and will summarise how these relate to study design. It is clear that both featural and holistic processes play a part in the construction of a composite. However, the relationship between these processes and the study design and measurement of composite quality employed has, in my view, been insufficiently explored. Second, I will summarise the relationship between three distinct areas related to composite research: practical utility of composites in police investigations; chosen experimental design; and police policy, or advice on best-practice. I will discuss the implications of the empirical investigations into police practice described in Chapter 5 for external validity in experiments. I will then discuss the contribution that experiments can make to the development of best practice. Third, I will explore in depth the concept of composite quality. In Chapter 2, I elaborated on the different definitions of composite quality. In Chapter 3 I considered the value and virtue of the different measurements of composite quality. Based on the issues discussed in these chapters, Study 5 found that composite quality is a kaleidoscope of ideas and measurements, rather than a single, unified entity. I will therefore go on to propose a more precise description of composite quality that takes into account its multiple facets. Finally, on the basis of what I have found in this thesis I will offer some directions for further work. My aim here is to provide a basis for future research and policy formulation, rather than to provide ready answers. I thereby hope to contribute to the construction of a theoretical and methodological basis for the domain of composite construction.

7.1 Featural and Holistic Processes in Composite Construction

In Chapter 2, I concluded that featural, configural as well as holistic processes are involved when we perceive and retrieve faces. The cardinal argument regarding the processes involved in face composite construction was that the influence of these three processes weigh differently in different cognitive tasks. In face recognition, we do not analyse separate features of the face, but the face as a whole. Therefore, more holistic processes will be involved while a witness compares the suspect in a line-up to her memory of the perpetrator. Similarly, recalling aspects of a face often makes it necessary to dissect the face into features and configurations.

With respect to the relationship between these three facial properties and composite systems, it has been argued that the new generation of eigenface-based software systems draws more heavily on the process of face recognition, and therefore on the stronger and more automatic retrieval of a holistic face impression. Critics of these systems argue that feature-assembly systems like Identikit and Pro-Fit by now have incorporated configural and holistic aspects, allowing witnesses to choose and change features within the context of a face. Supporters of the eigenface-based systems counter that witnesses using the older systems still have to choose individual features, thus focussing more on featural processes than on holistic ones. Investigating this controversy, Study 1 found that composites constructed with Identikit, a feature-based system, were more often correctly recognized, and were rated more similar to the target face than were composites constructed with ID, an Eigenface-based system. These results are in line with previous studies comparing both approaches. Previous studies have either found no significant difference or an advantage to feature-based systems like FACES or E-Fit, as compared with Eigenface-based systems like ID (Tredoux et al., 2007) and Evo-Fit (Frowd et al., 2005a, 2005b). However, it is worth noting that both ID and Evo-Fit were still under construction at the time that these studies were conducted, and that a target-absent line-up, employed only in the current study, showed a significant increase in correct rejections when the judgement was based on composites produced with a holistic instead of featural system. The conclusion that either approach is more successful than the other may therefore be premature.

Within composite research the varying influence of featural and holistic aspects during face processing is associated not only with retrieval but also with encoding of a face. Participant-witnesses, in contrast to real-life witnesses, know that they will subsequently be required to construct a composite. They might therefore be encouraged to try to analyse and disintegrate the face in an attempt to remember it better, thereby disrupting its integrity. All the experiments described above, including Study 1, utilized informed witnesses for composite construction. If one assumes that Wells and Hryciw (1984) are correct in assuming that a featural encoding strategy will encourage featural retrieval, then these results are no longer surprising. It is possible that participants in Study 1, and in those studies conducted by Frowd et al. (2005a, 2005b) and Tredoux et al. (2007), might have utilized a feature-based encoding strategy, and subsequently produced better-quality composites when a feature-based retrieval system was offered. It would be interesting to see if Eigenface-based systems can outperform feature-based systems when unsuspecting participant-witnesses are exposed to a live target, and only later instructed to construct a composite of this target.

In terms of these arguments, participants using a feature-based encoding strategy should construct better-quality composites when using a feature-based system and a featural context-reinstatement interview. Likewise, unintentional learning of the target face, when matched with an eigenface-based system and a holistic interview, should yield better quality composites than any other combination of factors. In Study 1, in which participants were presented with a video of the target and were informed that they would have to reconstruct it at a later stage, the use of Identikit as the feature-based system yielded more similar and more identifiable composites than did use of ID as the holistic system, but not more correct rejections of the target-absent line-ups. However, neither the featural context-reinstatement interview nor the holistic interview when applied in conjunction with a holistic system appeared to have any effect on composite quality. These results are in keeping with studies conducted by Frowd et al. (2005c, 2007a) and Koehn and Fisher (1997). With the exception of Davies and Milne (1985) and Frowd et al. (2008), none of the studies investigating context-reinstatement as an interview method found a consistent increase in composite quality.

Summarizing influences of featural or holistic processing of a face on composite quality, one could argue that an advantage of holistic retrieval aids could not be found. Holistic processing, described as the stronger, more successful cognitive process present during face recognition, is argued to be beneficial in face composite construction. Yet, the above mentioned studies have repeatedly found that holistic composite systems and holistic context-reinstatement do not achieve better-quality composites consistently. However, both encoding strategy and composite quality measurement might have contributed to these findings. As discussed in Chapter 2, it can be argued that if a face is encoded more featurally than holistically, a holistic retrieval strategy might be unable to contribute to a better composite. Relating these featural and holistic processes to the composite itself, two possible speculations can be made. On the one hand, it could be that these processes contribute to a generally better-quality composite. Alternatively, featural strategies might only increase composite quality with respect to featural properties, resulting in an improved score when measured with feature-based methods but not when measured with holistic ones. This distinction is important in that it determines how differences in the results of composite quality measurement can be explained. As discussed in Chapters 3 and 6, different properties or qualities might be inherent in each composite. One of these properties might be featural similarity or likeness, whereas another could be the similarity or likeness of the holistic impression. Thus, a measurement of the Euclidean distances within a face as applied by O'Toole (in press) and Rhodes (1988) can be described as a physical feature-and-configuration-measure. The measurement of holistic dimensions of a composite might be more difficult unless it is defined as the property mainly responsible for recognizing the composite as a specific person. The notion of featural and holistic processing might therefore not only be related, for example, to the system used, but also to other experimental design factors such as stimulus presentation during encoding, the interview method used while constructing the composite, and the composite quality measurement technique. A correlation between featural or holistic processing during retrieval and the resulting composite quality is difficult to find if one or both of these variables are ill-defined, as is the case with composite quality. These difficulties increase if one of these variables, particularly composite quality, is related to multiple aspects of the composite.

7.2 Relating Practical Utility to Experimental Design

Research on composite construction, as stated in Chapter 5, serves two overall purposes: to develop theoretical knowledge on how faces are processed, and to advise law-enforcement agencies on best-practice. Comparing police practice to experimental practice is a necessary step in the establishment of the external validity of experiments, and in the provision of advice on best-practice. Two studies, described in Chapter 5, have shown that it is possible, although challenging, to research face composite construction in the field.

To name but a few aspects, witnesses in South Africa often construct composites after a longer delay than is used in experiments, and under circumstances in which the suspect is of a different race and was encountered under stressful conditions in which the threat of physical harm to the witness was present. Composites are shown to informers of the police or put up on notice boards, yielding few helpful responses. Composites are mostly constructed in cases in which no other evidence is available. Further information would be required to determine if other similar cases, namely those involving contact crimes with a surviving witness, also have this lack of other evidence, or if composites are requested in only those cases in which there is no other evidence. An indicator here could be the time it took the investigating officer to request a composite construction. Investigating officers typically approached the composite-construction unit three days after the crime, two days later than requested by South African police policy. This raises the possibility that the latter interpretation might be true: composites are most often used as a last resort in an investigation in which no other evidence furthering the investigation or leading to a suspect is available. If this is indeed the case, the recommendation made by McNamara (2009) that composites should be constructed when necessary instead of when possible, has already been met.

Studies 3 and 4, in conjunction with studies by Kitson and Darnbrough (1978) and McQuiston et al. (2006), achieved a better understanding of real-life composite construction. Kitson and Darnbrough stated that study results could be used to improve the system Photo-Fit, at that stage a relatively new system that was still under development. McQuiston-Surrett et al.'s findings point out the lack of standardisation within composite construction, in the police officer's training as well as in the way a composite is constructed. Studies in Chapter 5 showed the discrepancy between how composites are constructed in the laboratory, how they are constructed in real-life, and how they should be constructed according to police policy.

With respect to the time delay between crime encounter and composite construction, 18 and 13 days in the two studies, respectively, these are far longer intervals than most laboratory studies employ and that policy demands. On the one hand, Study 2 showed that even after such an extended delay, composites enable recognition at higher than chance level. On the other hand, Deffenbacher et al. (2008) and Shapiro and Penrod (1986) showed that face recognition decreases considerably during this time. In relation to policy implementation, this might mean that police have to decide if they want to incorporate composites in their initial investigation, which would produce a shorter delay, or if they want to use them only as a last resort, which would produce a longer delay. If it is used as the latter, further studies might show that scepticism regarding composite construction leads to a self-fulfilling prophecy: if composites are constructed late in an investigation, witnesses' memories are likely to have deteriorated, resulting in a bad-quality composite, which will in turn not be useful for the investigation, thus confirming the initial scepticism.

Composite utility was estimated by enquiring about composite distribution and subsequent responses from either the general public or specific persons. Most composites were shown to informers or distributed to other police stations or in the community in which the crime took place. For example, composites might have been sent to neighbouring police stations to enquire if a similar-looking suspect, being investigated for a similar offence might be known there. Composites are also put up on notice boards that police officers pass on a daily basis, allowing them to see and memorize the composite face. If they then come across a suspect showing a resemblance to the composite face they might investigate that suspect further. Both examples involve comparison of a face in view with a face from memory. However, it is unlikely that a police officer will have photographs of suspects with him when he comes across the notice board displaying the relevant suspect, enabling him to make a direct comparison. Similarly, a police officer might not carry photocopies of composite faces with him when encountering a new suspect, rather judging from memory whether the suspect shows any similarity to a composite he has seen. It is far more likely that a comparison between the facial composite and the face of the suspect will not be made with both in view. Such comparisons rely on recognition of the currently perceived facial image as based on a memory of another facial image. This process is related to the measures of recognition used in judging composite quality. Further study would be required to investigate what would happen if composite or suspect were in view while recognition occurred. Further study of composite utility could also reveal if, as Frowd et al. (2005c) claim, recognition of a known person is the most prominent response to composite distribution.

The investigation into comparability between composite and suspect image regrettably did not succeed. If it were successful, an experimental comparison could have investigated composite quality in the field. One could have investigated composite likeness and recognizability with some of the measurement techniques described in Chapter 3, estimating diagnosticity of composites by designing measurement techniques with a known task difficulty. One could further have had participant-witnesses construct a composite from the suspect's photograph, and compared composite quality between composites constructed by real witnesses and by participant-witnesses in the laboratory. Such a comparison would produce insight into the effects of stress on memory for faces and into the effects of intentional learning on composite quality. However, such studies are unlikely to be successful in South Africa, where access to suspect photographs is denied and there is a low conviction rate.

7.3 The Kaleidoscope of Composite Quality

Just as a kaleidoscope contains mirrors and different coloured beads and pebbles, allowing the viewer to see different patterns when looking through the lens, so does the researcher see different outcomes of experiments through the lens of composite quality measurement. The researcher usually tries to analyse every aspect of these patterns, investigating why certain patterns of results occur. However, with every change in the experimental design this pattern of results seems to change. Focussing solely on the outcome, this subsection considers the aspects of composite quality making up the multitude of different composite study results.

I believe that composite quality, as measured in most experiments, is relative rather than absolute. A composite might have a certain level of recognizability; however, the recognition rate in an experiment is related to task difficulty as much as it is related to composite recognizability. A bald man might be recognized as the person depicted in the composite for a lack of alternatives. Highly similar foils in a line-up might all be the same type likeness as the target, thus making a distinction between them difficult solely on the basis of the composite. Likewise, similarity ratings of a composite-target combination might be difficult to judge if no similarity scale with pre-defined, meaningful upper and lower limits exists. In short, if composites are to be described with terms such as “good”, “accurate”, or “bad,” they need to have been evaluated with a technique with a pre-assessed task difficulty. If there is no estimate of task difficulty which defines upper and lower expected performance, the assessment of composite quality should be restricted to relative rather than absolute terms. In my opinion the only justified absolute judgement is that of usefulness in an investigation: if a composite leads the investigation to a certain suspect, who in turn is incriminated by other evidence and subsequently convicted of the crime in question, this composite has proved its usefulness.

A composite is defined as a visual representation of a face, derived from the personal impression stored in the memory of the witness and translated through the composite system and the police operator into an image. The quality of this pictoral representation of the target face is inherent in the displayed face, and has to be estimated through techniques which in turn influence or bias the result. The quality of a composite as the final output of an eyewitness testimony cannot be judged directly, but only through indirect measures utilizing the judgement of others.

Returning to the discussion of the dimensions of composite quality in Chapter 3, past research has measured several aspects of composite quality. Studies 5 and 1 added more aspects of measured composite quality to this group. As discussed earlier, composites are regarded as having featural (including configural) and holistic properties. Recognizability as a specific target is considered a further dimension of composite quality, with the target either known or unknown to the person judging the composite. Recognizability, in contrast to holistic aspects, is measurable through techniques relying on identification of a target. Subjective composite likeness is, like recognizability, a more measurable concept. Composites in an experiment can, for example, be described as being more recognizable as an unfamiliar person in a from-memory comparison. Alternatively, a composite can be described as being more similar to the target face than a foil face when making an in-view comparison.

Therefore, judging one composite as better than another may not be appropriate. Rather, composites might be better described as more or less recognizable as a familiar or unfamiliar person, with a higher or lower likeness to the target person, this judgement being made while neither, one or both images are in view.

Adding to these distinctions, I propose two more dimensions as being inherent in the composite itself: diagnosticity and willingness to choose. Composite quality as measured in Study 5 seems to measure two distinctly different components. Both concepts have been extensively defined and studied in eyewitness identification research. Regarding the composite as the manifested memory of the witness for the target face, it is possible to imagine why these dimensions might also influence composite quality measures. In Study 1 as well as in Study 5, one experimental condition seemed to outperform the other condition in achieving a better identification rate in the target-present line-up. However, with reference to results of the target-absent line-up, this seemingly better experimental condition also showed a higher rate of identifications, only in this case they were incorrect. The "better" experimental condition seemed to have produced a more recognizable composite, but also seemed to have lowered the decision criteria in such a way that the participant-judge was more willing to choose, independent of whether the target was present or not. This distinction between diagnosticity and willingness to choose can be made in tasks in which a target-absent condition exists, and in which the participant can reject the choice, either by not identifying any specific person or by attributing a low likeness rating.

Summarizing aspects of composite quality, any interpretation of the influence of cognitive changes in the composite construction process is difficult, and likely to be negatively affected if composite quality refers to various different aspects whose correlations have not yet been sufficiently researched. If composite quality is indeed a kaleidoscope of different dimensions, the correlation between those dimensions has to be known in order to make valid comparisons between studies relying on different dimensions. For example, it is logical to assume that recognizability and diagnosticity are in some way related, and that both rely on the recognition of the whole face rather than of its individual parts. However, recognizability in composite research is used to describe the identification of known or unknown persons, from memory or in view, mostly without controlling for willingness to choose. Diagnosticity, as the more elaborately defined concept adopted from eyewitness recognition research, can contribute a more precise definition of recognizability by providing tools to measure utility and task difficulty and by controlling for willingness to choose. Interestingly, both concepts of diagnosticity and willingness to choose can be used to interpret both recognition measures as well as likeness ratings.

7.4 Implications and Application for Future Research

Composite research, among other things, lacks clear definitions of cognitive concepts underlying composite construction and evaluation. It also lacks clear guidelines by which to judge if an experimental design is appropriate for evaluation of cognitive phenomena.

One of the aims of this thesis was to develop a clearer understanding of the concept of composite quality. It thus investigated cognitive as well as methodological aspects of quality measurement. Featural, configural and holistic aspects of facial composites are rarely measured in isolation. Composite aspects like recognizability and likeness, familiar and unfamiliar face identification, from-memory and in-view similarity, diagnosticity and willingness to choose seem to be the dimensions in which composites are distinguishable. I believe that comparison of composite quality measurement, and understanding of the contribution of the various dimensions thereof to the outcome of those quality measurements, would the increase understanding of what composite quality entails, and of how it can be measured. Further explorations into how composites are used in police investigations are needed to judge which of the quality measurements is the most appropriate in terms of forensic usefulness. For example, if the identification of neighbour Thabo is shown to be the most prominent successful result of composite publication in a police investigation, experiments could focus on measuring composite quality defined as the recognizability of a known person.

Finally, it is my opinion that the triangle of research, practice and policy should be more balanced. If research aims to design experiments relevant for police practice, sound investigation of this practice would be beneficial. Likewise, in order to assess the possibilities for advice on best practice yielded by laboratory research, it is necessary to determine the difference between laboratory research and current police practice.

In my opinion Deffenbacher (2008) is correct in stating that a sound theoretical analysis of the operations of each variable affecting eyewitness performance is necessary for reaching that point at which a research domain can claim to have knowledge about how a variable influences eyewitness performance. With more than 40 studies published, the domain of face composite construction research is approaching a stage in which the multiplicity of findings could cause the common threads to be lost. There might, however, be enough evidence from a variety of studies to combine findings into a theoretical framework. This thesis contributes some cognitive and methodological considerations on which such a framework could be based.

References

Barry, C., Johnston, R. A., & Scanlan, L. C. (1998). Are faces “special” objects? Associative and semantic priming of face and object recognition and naming. *Quarterly Journal of Experimental Psychology*, 51, 853-882.

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

Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law & Human Behavior*, 29, 279-301.

Bennett, P. (1986). Face recall: A police perspective. Human Learning, 5, 197-202.

Berman, G. L., & Cutler, B. L. (1998). The influence of processing instructions at encoding and retrieval on face recognition accuracy. *Psychology, Crime & Law*, 4, 89-106.

Bower, G. H., & Karlin M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, 103, 751-757.

Boylan, J. (2000). *Portraits of guilt: The woman who profiles the faces of America’s deadliest criminals*. New York: Pocket books.

Brace, N. A., Pike, G. E., Allen, P., & Kemp, R. I. (2006a). Identifying composites of famous faces: Investigating memory, language and system issues. *Psychology Crime & Law*, 12, 351-366.

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 : Traditional questions and new ideas* (pp. 272-276). Krakow, Poland: Institute of Forensic Research Publishers.

Brace, N. A., Pike, G. E., Kemp, R. I., Turner, J., & Bennett, P. (2006b). Does the presentation of multiple facial composites improve suspect identification? *Applied Cognitive Psychology 20*, 213-226.

Brown, C., & Lloyd-Jones, T. J. (2005). Verbal facilitation of face recognition. *Memory and Cognition*, 33, 1442-1456.

Brown, C., & Lloyd-Jones, T. J. (2006). Beneficial effects of verbalization and visual distinctiveness on remembering and knowing faces. *Memory and Cognition*, 34, 277-286.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, 73, 105-116.

Bruce, V. (1988). *Recognising Faces.* Hove, U.K.: Lawrence Erlbaum Associates.

Bruce, V. (2003). *Getting to know you – How we learn new faces.* Final report to the Economic and Social Research Council. London: ESRC.

Bruce, V., Burton, A. M., & Craw, I. (1992). Modeling Face Recognition. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, 335, 121-128.

Bruce, V., Ness, H., Hancock, P.J.B., Newman, C. and Rarity, J. (2002). Four heads are better than one: Combining face composites yields improvements in face likeness. *Journal of Applied Psychology*, 87, 894-902.

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

Bruce, V. & Young, A. (1998). *In the eye of the beholder: The science of face perception*. Oxford: Oxford University Press.

Burger, J. (2007). *Strategic perspectives on crime and policing in South Africa.* Pretoria: Van Schaik Publishers.

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 51, 256-284.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video. *Psychological Science*, 10, 243-248.

Casey, S. J., & Newell, F. N. (2007). Are representations of unfamiliar faces independent of encoding modality? *Neuropsychologia*, 45, 506-513.

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

Christie, D., Davies, G., Shepherd, J., & Ellis, H. (1981). Evaluating a new computer-based system for face recall. *Law & Human Behavior*, 5, 209-218.

Clark, C. (2000). *Interviewing for facial identification*. Report to the Home Office Police and Reducing Crime Unit. London: Home Office.

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law & Human Behavior*, 29, 575-604.

Clark, S., Howell, R. T., & Davey, S. L. (2008). Regularities in Eyewitness Identification. *Law & Human Behavior*, 32, 187-218.

Clifford, B. R., & Davies, G. M.(1989). Procedures for obtaining identification evidence. In D. Raskin (Ed.), *Psychological methods in criminal investigation and evidence.* (pp. 47-96). New York: Springer Verlag.

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology*, 17, 97-116.

Cooper, E. E., & Wojan, T. J. (2000). Differences in the Coding of Spatial Relations in Face Identification and Basic Level Objects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26, 470-488.

Cormack J. (1979). *The police artists' reference*. Pewaukee, WI: Waukesha County Technical Institute.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. Journal of Applied Psychology, 73, 281-290.

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

Cutler, B. L., Stocklein, C. J., & Penrod, S. D. (1988). Empirical examination of a computerised facial composite production system. *Forensic Reports*, 1, 207-218.

Dando, C., Wilcock, R., & Milne, R. (2009). The cognitive interview: the efficacy of a modified mental reinstatement of context procedure for frontline police investigators. *Applied Cognitive Psychology*, 23, 138-147.

Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a a problem-solving experiment. *Behavior Research Methods*, 40, 428-434.

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 : recent psychological research and new perspectives*. (pp. 103 - 124). Chichester: Wiley & Sons.

Davies, G. M. (1986b). Capturing likeness in eyewitness composites: The police artist and his rivals. *Medicine, Science and the Law*, 26, 283-290.

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., Ellis, H., & Shepherd, J. (1978a). Face identification. The influence of delay upon accuracy of Photofit construction. *Journal of Police Science and Administration*, 6, 35-42.

Davies G., Ellis, H., & Shepherd, J. (1978b). Face recognition accuracy as a function of mode of representation. Journal of Applied Psychology, 63, 180-187.

Davies, G., & Little, M. (1990). Drawing on memory: Exploring the expertise of the police artist. *Medicine, Science and the Law*, 30, 345-353.

Davies, G., & Milne, A. (1985). Eyewitness composite production. A function of mental or physical reinstatement of context. *Criminal Justice and Behavior*, 12, 209-222.

Davies, G., Milne, A., & Shepherd, J. (1983). Searching for operator skills in face composite reproduction. *Journal of Police Science and Administration*, 11, 405-409.

Davies, G., & Oldman, H. (1999). The impact of character attribution on composite production: A real world effect? *Current Psychology*, 18, 128-139.

Davies, G. M., Shepherd, J. W., & Ellis, H. D. (1979). Similarity effects in face recognition. *American Journal of Psychology*, 92, 507-523.

Davies, G., Shepherd, J. W., Shepherd, J., Flin, R., & Ellis, H. (1986). Training skills in police Photofit operators. *Policing*, 2, 35-46.

Davies, G. M., & Valentine, T. (2006). Facial composites: Forensic utility and psychological research. In R. C. L. Lindsay, D. F. Ross, & J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology Vol. 2* (pp. 59–86). Mahwah, NJ: Erlbaum.

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.

Davis, M. R., McMahon, M., & Greenwood, K. M. (2005). The efficacy of mnemonic components of the cognitive interview: Towards a shortened variant for time-critical investigations. *Applied Cognitive Psychology*, 19, 75-93.

Deffenbacher, K. A. (1986). On the memorability of the human face. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), Aspects of face processing (pp. 61-70). Dordrecht, Netherlands: Martinus Nijhoff.

Deffenbacher, K.A. (2008). Estimating the impact of estimator variables on eyewitness identification: A fruitful marriage of practical problem solving and psychological theorizing. *Applied Cognitive Psychology*, 22, 815-826.

Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology-Applied*, 14, 139-150.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law & Human Behavior*, 28, 687-706.

Dekle, D. J. (2006). Viewing composite sketches: Lineups and showups compared. *Applied Cognitive Psychology*, 20, 383-395.

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

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.

Domingo, F. (1984, June). Composite art: The need for standardization. *Identification News*, 7-15.

Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, 111, 524-542.

Ebbesen, E. B., Flowe, H. D. (2002). Simultaneous v. Sequential Lineups: What Do We Really Know? Unpublished manuscript, Department of Psychology, University of California, San Diego, United States of America.

Ellis, A. W. (1992). Cognitive mechanisms of face processing. *Philosophical Transactions of the Royal Society of London, Series B*, 335, 113-119.

Ellis, H. D. (1984). Practical aspects of face memory. In G. Wells, & E. Loftus (Eds.), *Eyewitness testimony: psychological perspectives* (pp. 12-37). Cambridge: Cambridge University Press.

Ellis, H. (1986). Face recall: a psychological perspective. *Human Learning*, 5, 189-196.

Ellis, H. D. (1990). Developmental trends in face recognition. *The Psychologist: Bulletin of the British Psychological Society* , 3, 114–119.

Ellis, H., Davies, G., & McMurran, M. (1979). Recall of white and black faces by white and black witnesses using the Photofit system. *Human Factors,* 21, 55-59.

Ellis, H., Davies, G., & Shepherd, J. (1976). *An investigation of the Photofit system for recalling faces.* Swindon: Social Science Research Council.

Ellis, H., Davies, G., & Shepherd, J. (1978). A critical examination of the Photofit system for recalling faces. *Ergonomics,* 21, 297-307.

Ellis, H. D., & Shepherd, J. W. (1992). Face memory: Theory and practice. In P. Morris, M. Gruneberg, & R. Sykes (Eds.), *Aspects of memory, Vol. 1* (pp. 136-143). London: Routledge.

Ellis, H., Shepherd, J., & Davies, G. (1975). An investigation of the use of the Photofit system for recalling faces. *British Journal of Psychology*, 66, 29-37.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration*, 8, 101-106.

Ellis, H. D., & Young A. W. (1990). Accounting for Delusional Misidentifications. *The British Journal of Psychiatry*, 157, 239-248.

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

Fisher, R. P., & Quigley, K. L. (1992). Applying cognitive theory in public health investigations: Enhancing food recall with the cognitive interview. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 154-169). New York: Russell Sage Foundation.

Flowe, H. D., & Ebbesen, E. B. (2007). The effect of lineup member similarity on recognition accuracy in simultaneous and sequential lineups. *Law and Human Behavior*, 31, 33-52.

Frowd, C. D., Bruce, V., McIntyre, A., 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. (2007a). Parallel approaches to composite production: interfaces that behave contrary to expectation. *Ergonomics*, 50, 562-585.

Frowd, C. D., Bruce, V., Smith, A. J., & Hancock, P. J. B. (2008). Improving the quality of facial composites using a holistic cognitive interview. J*ournal of Experimental Psychology – Applied*, 14, 276-287.

Frowd, C. D., Carson, D., Ness, H., McQuiston-Surrett, D., Richardson, J., Baldwin, H., & Hancock, P. (2005a). 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. (2005b). A forensically valid comparison of facial composite systems. *Psychology Crime & Law*, 11, 33-52.

Frowd, C. D., Hancock, P. J., & Carson, D. (2004),. EvoFIT: A Holistic, Evolutionary Facial Imaging Technique for Creating Composites, ACM *Transactions of Applied Perceptions*, 1, 19-39.

Frowd, C. D., McQuiston-Surrett, D., Kirkland, I. & Hancock, P. J. B. (2005c). 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.

Frowd, C. D., McQuiston-Surrett, D, Anandaciva, S., Ireland, C. E., & Hancock, P. J. B. (2007b). An evaluation of US systems for facial composite production, *Ergonomics*, 50, 1987-1998.

Gauthier, I., Behrmann, M., & Tarr, M. J. (1999a). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience*, 11, 349-370.

Gauthier, I., & Tarr, M. J. (1997). Becoming a “greeble” expert: Exploring mechanisms for face recognition. *Vision Research*, 37, 1673-1682.

Gauthier, I.,Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999b). Activation of the middle fusiform ‘face area’ increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2, 568-573.

Geiselman, R. E., Fisher, R. P., Firstenberg, I., Hutton, L. A., Sullivan, S. J., Avetissian, I. V., & Prosk, A. L. (1984). Enhancement of eyewitness memory: An empirical evaluation of a cognitive interview. *Journal of Police Science and Administration*, 12, 74-80.

George, B., Gibson, S. J., Maylin, M. I., & Solomon, C. J. (2008). EFIT-V: interactive evolutionary strategy for the construction of photo-realistic facial composites. Proceedings of the 10th annual conference on Genetic and evolutionary computation, 1485-1490. Retrieved January 12, 2009, from http://delivery.acm.org/10.1145/1390000/1389384/p1485-george.pdf?key1=1389384&key2=8807864621&coll=GUIDE&dl=GUIDE&CFID=75299816&CFTOKEN=15350801

Gibling, F., & Bennett, P. (1994). Artistic enhancement in the production of Photofit likenesses – An examination of its effectiveness in leading to suspect identification. *Psychology, Crime & Law*, 1, 93-100.

Gibson, S. J., Solomon, C. J., Maylin, M. I. S., & Clark, C. (2009). New methodology in facial composite construction: from theory to practice. I*nternational Journal of Electronic Security and Digital Forensics*, 2, 156-168.

Gibson, S.J., Solomon, C. J., & Pallares Bejarano, A. (2003). Synthesis of Photographic Quality Facial Composites using Evolutionary Algorithms. In R. Harvey & J. A. Bangham (Eds.), *Proceedings of the British Machine Vision Conference 2003*, 1, 221-230.

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.

Green, D. M., & Swets J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Gwyer, P., & Clifford, B. R. (1997). The effects of the cognitive interview on recall, identification, confidence and the confidence/accuracy relationship. *Applied Cognitive Psychology*, 11, 121-145.

Halford, P., Milne, R., & Bull, R. (2005) *The identification performance of forensic eyewitnesses exposed to weapons and violence*. Paper presented at the XVth European Conference on Psychology and Law. Vilnius, Lithuania.

Hancock, P. J. B. (2000). Evolving faces from principal components. *Behavior Research methods, Instruments and Computers*, 32, 327-333.

Haxby, J. V., Hoffman, E.A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4, 223-233.

Hay, D. C., & Young, A. W. (1982). The human face. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 173-202). London: Academic Press.

Hinkle, D. P. (1990). *Mug shots: A police artist's guide to remembering faces.* Boulder: Paladin Press.

Heerwegh, D. (2006). An investigation of the effect of lotteries on Web survey response rates. *Field Methods*, 18, 205-220.

Jewkes, R., Sikweyiya, Y., Morrell R., & Dunkle K. (2009). Understanding men's health and use of violence. MRC Policy brief. Retrieved November 30, 2009, from http://www.mrc.ac.za/gender/violence_hiv.pdf

Johnston, R., & Edmonds, A. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, 17, 577-596.

Jones, T. C., & Bartlett, J. C. (2009). When false recognition is out of control: The case of facial conjunctions. *Memory & Cognition*, 37, 143-157.

Kitson, A., Darnbrough, M., & Shields, E. (1978). Let's face it. *Police Research Bulletin*, 30, 7-13.

Koehn, C., & Fisher, R. P. (1997). Constructing facial composites with the Mac-a-Mug Pro system. Psychology, *Crime & Law*, 3, 209-218.

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.

Koehnken, G., Milne, R., Memon, A., & Bull, R. (1994). A meta-analysis of the effects of the Cognitive Interview. Paper presented at the Biennial Conference of the American Psychology Law Society, Sante Fe.

Koehnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology Crime & Law*, 5, 3-27.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Towards a psychology of memory accuracy. *Annual Review of Psychology*, 51, 481-537.

Kovera, M. B., Penrod, S., Pappas, C., & Thill, D. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology*, 82, 235-246.

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology*, 49, 58-69.

Kuritz, S. J., Landis, J. R., & Koch, G. G. (1988). A general overview of Mantel-Haenszel methods: applications and recent developments. *Annual Review of Public Health*, 9, 123-160.

Laughery, K., Duval, C., & Wogalter, M. (1986). Dynamics of face recall. In H. Ellis, M. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 373-387). Dordrecht, The Netherlands: Nijhoff.

Laughery, K. R., Fessler, P. K., Lenorovitz, D. R., & Yoblick, D. A. (1974). Time delay and similarity effects in facial recognition. *Journal of Applied Psychology*, 59, 490–496.

Laughery, K., & Fowler, R. (1980). Sketch artist and Identi-Kit procedures for recalling faces. *Journal of Applied Psychology*, 65, 307-316.

Laughery, K. R., Smith, V. L., & Yount, M. B. (1980). Visual support devices: evaluation of a new technique for constructing facial images. Proceedings of the Human Factors Society, 24th Annual Meeting, New York, 1-4, 302-305. Retrieved March 23, 2008 from http://www.ingentaconnect.com/content/hfes/hfproc/1980/00000024/00000004/art00016

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology Section A – Human Experimental Psychology*, 53, 513-536.

Lee, K. J., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness and identification: Testing the face-space framework. *Psychological Science*, 11, 379-385.

Levi, A. M. (1998). Are defendants guilty if they were chosen in a lineup? *Law & Human Behavior*, 22, 389-407.

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

Liu, C. H., & Chaudhuri, A. (2000). Recognition of unfamiliar faces: three kinds of effects. *Trends in Cognitive Sciences*, 4, 445-446.

Lund, N. (2001). *Attention and pattern recognition*. Hove, U.K.: Routledge.s

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.

Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distractors for lineups. *Law and Human Behavior*, 15, 43-57.

Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law & Human Behavior*, 5, 299–309.

Malpass, R. S. (1993). "They all look alike to me." In M. Merrens, & G. Brannigan (Eds.), *The undaunted psychologist* (pp. 74-88). New York: McGraw-Hill.

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

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, 4, 482-489.

Malpass, R. S., & Lindsay, R. C. L. (1999). Measuring line-up fairness. A*pplied Cognitive Psychology*, 13, 1-8.

Malpass, R. S., Tredoux, C. G., Compo, N. S., McQuiston-Surrett, D., Maclin, O. H., Zimmerman, L. A., Topp, L.D. (2008). Study space analysis for policy development. *Applied Cognitive Psychology*, 22, 789-801.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002) The many faces of configural processing. *Trends in Cognitive Sciences*, 6, 255-260.

McNamara, J. M. (2009). Sketchy Eyewitness-Identification procedures: A proposal to draw up legal guidelines for the use of facial composites in criminal investigations. *Wisconsin Law Review*, 3, 763-799.

McNeil, J. E., Wray, J. L., Hibler, N. S., Foster, W. D., Rhyne, C. E., & Thibault, R. (1987). Hypnosis and Iidenti-kit: A study to determine the effect of using hypnosis in conjunction with the making of Identi-kit composites. *Journal of Police Science and Administration*, 15, 63-67.

McQuiston-Surrett, D., Topp, L. D., & Malpass, R. S. (2006). Use of facial composite systems in US law enforcement agencies. *Psychology Crime & Law*, 12, 505-517.

McQuiston-Surrett, D., & Topp, L. D. (2008). Externalizing Visual Images: Examining the Accuracy of Facial Descriptions vs. Composites as a Function of the Own-Race Bias. *Experimental Psychology*, 55, 195-202.

Megreya, A. M., & Burton, A. M. (2008). Matching Faces to Photographs: Poor Performance in Eyewitness Memory (Without the Memory). *Journal of Experimental Psychology-Applied*, 14, 364-372.

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 bias in memory for faces – A meta-analytic review. *Psychology*, Public Policy and Law, 7, 3-35.

Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for own- and other-race faces: A dual-process approach. *Applied Cognitive Psychology,* 19, 545-567.

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

Meissner, C. A., Sporer, S. L., & Susa, K. J. (2008). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *European Journal of Cognitive Psychology*, 20, 414-455.

Memon, A., & Higham, P. (1999). A review of the cognitive interview. Psychology, *Crime and Law*, 5, 177-196.

Municipal Demarcation Board (2001). Population distribution and dynamics. Retrieved 19 October, 2007, from http://www.demarcation.org.za/

National commissioner of the South African Police Service (2005). Annual Report of the South African Police for 2004/2005. Retrieved 12 February, 2006, from http://www.saps.gov.za/statistics/reports/crimestats/2005/crime_stats.htm

National commissioner of the South African Police Service (2006). Annual Report of the South African Police for 2005/2006. Retrieved 30 March, 2007, from http://www.saps.gov.za/statistics/reports/crimestats/2006/crime_stats.htm

National commissioner of the South African Police Service (2007). Annual Report of the South African Police for 2006/2007. Retrieved 2 May, 2008, from http://www.saps.gov.za/statistics/reports/crimestats/2007/crime_stats.htm

O'Neil, K. M., Penrod, S. D., & Bornstein, B. H. (2003). Web-based research: Methodological variables' effects on dropout and sample characteristics. *Behavior Research Methods Instruments & Computers*, 35, 217-226.

O'Toole, A. J. (2005). Psychological and neural perspectives on human face recognition. In S. Z. Li, A. Jain (Eds.), *Handbook of Face Recognition* (pp. 349-369). New York: Springer-Verlag.

O'Toole, A. J. (in press). Cognitive and Computational Approaches to Face Perception. In A. Calder, G. Rhodes, J. V. Haxby, & M. Johnson (Eds.), *Handbook of Face Perception*. Oxford: Oxford University Press.

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

Penry, J. (1971). *Looking at faces and remembering them: A guide to facial identification*. London: Elek Books.

Poole, O. (2004, January 21). I know what it's like to want justice. *The Daily Telegraph*, p.14.

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. *Journal of Applied Psychology*, 84, 167-176.

Rakover, S. (2002). Featural vs. configurational information in faces: A conceptual and empirical analysis. *British Journal of Psychology*, 93, 1-30.

Rakover, S. S., & Cahlon, B. (2001). Face Recognition: *Cognitive and Computational Process*. Amsterdam: John Benjamins Publishing Company.

Rhodes, G. (1988). Looking at Faces: First-order and Scond-order Features as Determinants of Facial Appearance. *Perception*, 17, 43-63.

Schmidt, H. C., & Frowd, C. .F. (2006). An investigation into the cross-race effect in face composite construction. Proceedings of the 16th Conference of the European Association of Psychology and Law, 67-76. Retrieved 29 January, 2010, from http://www.heikeschmidt.com/cross-race.pdf

Schmidt, H. C., & Tredoux, C. G. (2006). Utilisation and usefulness of face composites in the South African Police Service – an evaluation study. *South African Journal of Criminal Justice*, 19, 303-314.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology*, 22, 36-71.

Schwarzer, G. (2000). Development of face processing: The effect of face inversion. *Child Development*, 71, 391-401.

Shapiro, P. N., & Penrod, S. (1986). A Meta-analysis of facial identification studies. *Psychological Bulletin*, 100, 139-156.

Sirovich, L., & Kirby, M. (1987). Low dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4, 519-524.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology-Human Learning and Memory*, 5, 460-471.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental Context and Human Memory. *Memory & Cognition*, 6, 342-353.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, 8, 203-220.

South African Police Service (2003). *Policy regarding the Facial Identification section*. Unpublished manuscript, South African Police Service, Cape Town: SAPS

Sporer, S. L. (1989). Alternative Ansätze in der Rechtspsychologie. [Alternative approaches in psychology and law]. *Zeitschrift für Sozialpsychologie*, 20, 259-265.

Sporer, S. L. (1993a). Clothing as a contextual cue in facial recognition. German *Journal of Psychology*, 17, 183-199.

Sporer, S. L. (2001). The cross-race effect – Beyond recognition of faces in the laboratory. *Psychology Public Policy and Law*, 7, 170-200.

Sporer, S. L. (2007). Person descriptions as retrieval cues: Do they really help? *Psychology Crime & Law*, 13, 591-609.

Steblay, N. M. (1992). A Meta-analytic review of the weapon focus effect. *Law & Human Behavior*, 16, 413-424.

Steblay, N. K. (2008). Commentary on "Studying Eyewitness Investigations in the Field": A look forward. *Law & Human Behavior*, 32, 11-15.

Stephan, C. N., & Cicolini J. (2008). Measuring the Accuracy of Facial Approximations: A Comparative Study of Resemblance Rating and Face Array Methods. *Journal of Forensic Sciences*, 53, 58-64.

Tanaka, J. W., & Farah, M. J. (1993). Parts and Wholes in Face Recognition. The Quarterly Journal of Experimental Psychology: *Human Experimental Psychology*, 46A, 225-245.

Tanaka, J.W., & Farah, M. J. (2003). The holistic representation of faces. In: M. A. Peterson, & G. Rhodes, *Perception of faces, objects and scenes* (pp. 53-74). Oxford: Oxford University Press.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, 25, 583-592.

Taylor, K. T. (2001). Forensic art and illustration. Boca Raton, FL: CRC Press.

Thomson, D. M., Robertson, S. L., & Vogt, R. (1982). Person recognition: the effect of context. *Human Learning*, 1, 137-154.

Tollestrup, P., Turtle, J., & Yuille, J. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144-160). New York: Cambridge University Press.

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law & Human Behavior*, 22, 217-237.

Tredoux, C. (2001). *A non-verbal approach to face reconstruction by eyewitnesses*. Technical Report 15/1/3/16/0158, Centre for Science Development, Human Science Research Council, South Africa.

Tredoux, C., Nunez, D., Oxtoby, O., & Prag, B. (2007). An evaluation of ID: an eigenface based construction system. *South African Computer Journal*, 37, 90-96.

Tredoux, C., Rosenthal, Y., Nunez D., & da Costa, L. (1999). Face reconstruction using a configural, eigenface-based composite system. Paper presented to the third meeting of the Society for Applied Research Memory and Cognition, Boulder, Colorado. Retrieved December 5, 2003, from http://web.uct.ac.za/depts/psychology/plato/boulder.pdf

Tsao, D. Y., & Livingstone, M. S. (2008) Mechanisms of face perception. *Annual Review of Neuroscience*, 31, 411-437.

Tulving, E. (1985). How many memory systems are there? *The American Psychologist*, 40, 385-398.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247, 301-306.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.

Turtle, J., Read, J. D., Lindsay, D. S., & Brimacombe, C. A. E. (2008). Toward a more informative psychological science of eyewitness evidence. *Applied Cognitive Psychology*, 22, 769-778.

US Census Bureau (2010, January 7th) World POPClock Projection. Retrieved January 7, 2010 from http://www.census.gov/ipc/www/popclockworld.html

Valentine, T. (2001). Face-space models of face recognition. In: M. J. Wenger, & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 83-113). Mahwah, NJ: LEA.

Valentine, T., Darling, S., & Memon, A. (2007). How can psychological science enhance the effectiveness of identification procedures? An international comparison. *Public Interest Law Reporter*, 11, 21-39.

Valentine, T., & Mesout, J. (2009). Eyewitness Identification Under Stress in the London Dungeon. *Applied Cognitive Psychology*, 23, 151-161.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology*, 17, 969-993.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57, 137-154.

Wells, G. L. (1978). Applied Eyewitness-Testimony Research – System Variables and Estimator Variables. *Journal of Personality and Social Psychology*, 36, 1546-1557.

Wells, G. L. (2008). Field experiments on eyewitness identification: Towards a better understanding of pitfalls and prospects. *Law & Human Behavior*, 32, 6-10.

Wells, G. L., Charman, S. D., & Olson, E. A. (2005). Building face composites can harm lineup identification performance. *Journal of Experimental Psychology-Applied,* 11, 147-157.

Wells, G., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory and Cognition*, 12, 338-344.

Wells, G. L., Leippe, M. R., & Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law & Human Behavior*, 3, 285-293.

Wells, G. L., & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, 3, 776-784.

Wells, G. L., Malpass, R. S., Lindsay, R. C., Fisher, R. P., Turtle, J. W., & Fulero, S. M. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist*, 55, 581-598.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, 54, 277-295.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S.M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for line-ups and photo-spreads. *Law & Human Behavior*, 22, 603-647.

Wells, G. L., & Turtle, J. W. (1988). What is the best way to encode faces? In M. M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory: current research and issues, Vol. 1: Memory in everyday life* (pp. 163-168). Oxford: Wiley.

Wogalter, M., & Marwitz, D. (1991). Face composite construction – In-view and from-memory quality and improvement with practice. *Ergonomics*, 34, 459-468.

Woodhead, M. M., & Baddeley, A. D. (1981). Individual-Differences and memory for faces, pictures and words. *Memory & Cognition*, 9, 368-370.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology Public Policy and Law*, 7, 119-133.

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology*, 10, 75-84.

Wright, D. B., Memon, A., Skagerberg, E. M., & Gabbert, F. (2009). When Eyewitnesses Talk. *Current Directions in Psychological Science*, 18, 174-178.

Wright, D. B., & Stroud, J. N. (2002). Age Differences in Lineup Identification Accuracy: People Are Better with Their Own Age. *Law & Human Behavior*, 26, 641-654.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441-517.

Young, A. W., & Bruce, V. (1991). Perceptual Categories and the Computation of Grandmother. *European Journal of Cognitive Psychology*, 3, 5-49.

Young, A. W., & Burton, A. M. (1999) Simulating face recognition: Implications for modelling cognition. *Cognitive Neuropsychology*, 16, 1-48.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching Familiar and Unfamiliar Faces on Internal and External Features. *Perception*, 14, 737-746.

Yuille, J. C. (1993). We must study forensic eyewitnesses to know about them. *American Psychologist*, 48, 572-573.

Appendices

Appendix P-01: Study 2 – Scripted event

Appendix P-02: Study 3 – Survey form for the LCRC

Appendix P-03: Study 3 – Survey form for the investigating officer

Appendix P-04: Study 4 – Survey form for the LCRC ('LCRC form')

Appendix P-05: Study 4 – Survey form for the investigating officer ('IO form')

Appendix P-06: Study 4 – Information gathered from the LCRC database and dockets

Appendix P-07: Study 5 – Matching task answer sheet

Appendix P-08: Study 5 – Mugshot identification task answer sheet

Appendix P-09: Study 5 – Mugshot reduction task answer sheet

Appendix P-10: Study 5 – Naming task answer sheet

Appendix E-01: Study 1 – Composites and Target Photographs

Appendix E-02: Study 1 – Target-Absent and Target-Present Photo-Arrays

Appendix E-03: Study 1 – Video Target 1

Appendix E-04: Study 1 – Video Target 2

Appendix E-05: Study 1 – Video Target 3

Appendix E-06: Study 2 – Composites and Target Photographs

Appendix E-07: Study 2 – Target-Present Photo-Arrays

Appendix E-08: Study 5 – Composites and Target Photographs

Appendix E-09: Study 5 – Target and related foil photographs

Appendix P-01

Study 2 – Scripted Event

The following is a guideline of the mock event that participant-witnesses were exposed to. It is formulated in phrases rather than sentences to make it easier to follow. Nevertheless, these phrases, expression of emotions and actions were followed in each mock event with each target person (= 'Peter').

Short-term memory test in progress for approximately 5 minutes. Room is occupied by eight participant-witnesses and the experimenter (= Heike). All participants saw the experimenter put a sign on the door "Experiment in progress – do not enter") before closing it and starting the short-term memory test.

Hard knock on the closed door. Heike hesitates for a moment, and then continues with test. Peter enters.

Heike:	..sign at the door..
Peter:	Sorry.. my name is Peter … Secretary said you have the stopwatch and the calculator ... I need it now … Can I have it?
Heike:	… Experiment in progress ... you are disturbing ... come back later..
Peter :	..also have an experiment ... need that stuff … (*urgent*)
Heike:	Can't that wait? …be finished in 15 minutes (*getting angry*)
Peter:	(getting angry too) I also have experiment ... computer science department ... 50 people waiting for me ... 5 minutes (*looks around, sees the stopwatch in Heike's hand, points at it*) …that's the stopwatch ... where is the calculator? (*checking out room, looking at participants briefly*)
Heike:	…somewhere in this office
Peter:	…will be done in a second…(*walking around, checking where calculator is, specifically at desks behind participants, mumbling and apologetic to participants, eye-contact to each one, finding it on a computer, coming back to Heike, approaching her again*)
Peter:	Do you really need the stopwatch?
Heike:	Well, actually yes ... part of experiment…
Peter:	But I need it…(*towards audience, looks at them briefly*) Does anybody have a stopwatch or watch to give to her?

Either somebody in audience says he/she can give, or Heike says that she can use her wristwatch

Peter:	…sorry to interrupt (*and leaves*)

Appendix P-02

Archival Study – Internal Form

LCRC Case Number	
Investigating police station	
Police station case number	

LCRC Information

Date crime reported	
Date case received	
Date composite compiled	
Type of crime	
LCRC member	

Number of witnesses	
Number of suspects	

Per compiled suspect

Description of suspect	
Subjective resemblance	
Additional comments	

SAPS Information (to be completed when cased closed and conviction reached)

Conviction	Yes No Info not accessible
Confession	Yes No Info not accessible
Photo of suspect available	Yes No (if yes, photo number:)
Race of witness?	
Distribution of composite	Written press TV Informers Distribution in crime district None Police station notice board other station Showing to Witness before IDparade
Response on distribution	
Evidence other than composite	Fingerprints Idparade with same witness Idparade with different witness Surveillance camera Other:
Composite presented in court	Yes No Info not accessible

Appendix P-03

Facial Compilations – Cases processed by the LCRC in 2002

Police Station: XY

In the following table all cases are listed in which the above stated police station asked for the assistance of the LCRC in compiling an Identikit in the year 2002. Following up these cases we would like to know if the dockets have been closed and if a conviction of the suspect was reached. Please indicate the state of affairs in each of the following cases by ticking the correct answer alternative. After completing this list please fax it back to the University of Cape Town, for attention of Heike Schmidt, fax no. 021-650 4104. Your participation will help us a lot in evaluating the quality and further use of the produced identikits. Thanks you!

Case Number	**Closed docket?**	**Conviction reached?**
xx / mm / 2002	yes no	yes no
xx / mm / 2002	yes no	yes no
xx / mm / 2002	yes no	yes no
xx / mm / 2002	yes no	yes no
xx / mm / 2002	yes no	yes no
xx / mm / 2002	yes no	yes no
xx / mm / 2002	yes no	yes no
xx / mm / 2002	yes no	yes no
xx / mm / 2002	yes no	yes no
xx / mm / 2002	yes no	yes no

Appendix P-04

Evaluation of face composites – Form for LCRC

Investigating police station | Police station case number

Investigating officer | LCRC Case Number

Instruction

Please complete this form directly after constructing the face composite with the witness. If there is more than one witness, allocate a number to each of them and answer each question by specifying for which witness this is true. Attach the three pages of the *evaluation form for the police station* to the composite you send out.

Witness particulars:

Race of witness ☐ black ☐ coloured ☐ white

Gender of witness ☐ male ☐ female

Age of witness ____

Suspect's particulars: (only suspects whose composites were compiled)

Race of first suspect ☐ black ☐ coloured ☐ white

Gender of first suspect ☐ male ☐ female

Approximate age of first suspect ____

Race of second suspect ☐ black ☐ coloured ☐white

Gender of second suspect ☐ male ☐ female

Approximate age of second suspect ____

Crime particulars:

Type of crime: ☐ armed robbery ☐ robbery ☐ rape ☐ other: __________

Was a weapon used? ☐ yes ☐ no

Witness' involvement: ☐ victim ☐ bystander/observer

Was the witness physically assaulted? ☐ yes ☐ no

Approximate exposure time to suspect: ____ minutes How many suspects were involved in the crime? ____

Session particulars:

Length of session with witness: ___ minutes

Did the witness show any form of emotional distress in the interview? ☐ yes ☐ no

Specify form of emotional distress: ____

Which aspects of the interview seemed to have helped the witness in the reconstruction of the face?

____

____

Face composite evaluation –research by LCRC / University of Cape Town Page 1 of 4

Appendix P-05

Evaluation of face composites – Form for police station

Investigating police station | Police station case number

Investigating officer | LCRC Case Number

Instruction

Please indicate in *Part A* – Distribution - what the received face composite was used for and if it yielded any response. Specifically indicate if this response led to a person linked to the crime and suspected of committing this crime. If there is one or more than one suspect, allocate a number to each of them and answer each question by specifying for which suspect this is true.

Fill out *Part B and C* after the case is closed or else within 5 months after receiving the composite. Indicate in Part C if other evidence is available for this case. If a person was linked to the crime and suspected of committing this crime, indicate in *Part C* which of the other evidence supports the view that this suspect might be the perpetrator (short note next to the tick-of-box, or alternative unambiguous comment at the bottom of the page). If there is more than one suspect, proceed as proposed above, but assign a number to each suspect. If no suspect was found indicate available evidence nevertheless.

Please *return this form after the case is closed or else within 5 months after receiving the composite,* via fax to the University of Cape Town (021-6504104) or via fax/post to the LCRC Cape Town.

Thank you for helping us evaluate the use and usefulness of face composites produced in the LCRC. Your participation is greatly appreciated!

____________________ Capt. van der Westhuizen, LCRC

_________________________ Heike Schmidt, University of Cape Town

Part A - Distribution

After receiving the composite, was it…..

- **Published in the newspaper?** □ **yes** □ **no**

  With which other information was it distributed? □ Date and Location of crime
  □ Modus operandi
  □ Description of suspect □ Other: ____________

  Did this form of distribution lead to any response? □ yes □ no

  If yes, was this response helpful? □ yes □ no

  If yes, in which way was the response helpful? (please describe in short)

____

____

- **Shown on TV?** □ **yes** □ **no**

  With which other information was it distributed? □ Date and Location of crime
  □ Modus operandi
  □ Description of suspect □ Other: ____________

  Did this form of distribution lead to any response? □ yes □ no

  If yes, was this response helpful? □ yes □ no

  If yes, in which way was the response helpful? (please describe in short)

____

- **Shown to informers or specific people in the community?** ☐ **yes** ☐ **no**

With which other information was it distributed? ☐ Date and Location of crime
☐ Modus operandi
☐ Description of suspect ☐ Other: ____________

Did this form of distribution lead to any response? ☐ yes ☐ no

If yes, was this response helpful? ☐ yes ☐ no

If yes, in which way was the response helpful? (please describe in short)

____

____

- **Distributed to other police stations?** ☐ **yes** ☐ **no**

With which other information was it distributed? ☐ Date and Location of crime
☐ Modus operandi
☐ Description of suspect ☐ Other: ____________

Did this form of distribution lead to any response? ☐ yes ☐ no

If yes, was this response helpful? ☐ yes ☐ no

If yes, in which way was the response helpful? (please describe in short)

____

____

- **Distributed in the community where the crime took place?** ☐ **yes** ☐ **no**

With which other information was it distributed? ☐ Date and Location of crime
☐ Modus operandi
☐ Description of suspect ☐ Other: ____________

Did this form of distribution lead to any response? ☐ yes ☐ no

If yes, was this response helpful? ☐ yes ☐ no

If yes, in which way was the response helpful? (please describe in short)

____

____

- **Other form of distribution?** ☐ **yes** ☐ **no**

Please describe

____

____

Part B – Legal status

Five months after receiving the composite …

- the case is □ under investigation □ withdrawn □ undetected □ conviction
- a suspect has been □ interviewed □ arrested (including being on probation for this case) □ identified but wanted □ no suspect □ other: ________

Part C – Other evidence

- **Fingerprint evidence** □ **yes** □ **no**
  Fingerprints of the suspect found either at the scene of the crime or on property that had been taken from the scene or used to commit a crime. Established from evidence of the witnesses or from the suspect themselves that suspect never had legal access to those premises or that property on a previous occasion, and that their fingerprints could not be there for any innocent reason.
- **Property found in the possession of the suspect** □ **yes** □ **no**
  This includes not only property of the complainant that had been stolen but includes property used in the commission of the crime, for example a weapon that was proven to have been used to cause injury, or a tool that had been used to gain entry to premises.
- **Video evidence** □ **yes** □ **no**
  Surveillance cameras positioned at the scene of the crime. Police officers could e.g. identify suspects they knew from previous occasions.
- **Body sample** □ **yes** □ **no**
  Body sample defined as hair, blood type or DNA, either of the offender found on the victim or of the victim found on the offender or his/her property that could not be there for any innocent reason.
- **Suspect confession** □ **yes** □ **no**
  Suspect confessed to a commissioned police officer, a magistrate or in court.
- **Evidence of co-accused** □ **yes** □ **no**
  Co-accused confession of crime to either a commissioned police officer, a magistrate or in court including stating that the suspect was involved.
- **Identification** □ **yes** □ **no**
  Live or photo identification parade.
- **Other substantial evidence** □ **yes** □ **no**

  Please specify ____

Face composite evaluation –research by LCRC / University of Cape Town Page 4 of 4

Appendix P-06

Collaboration Study between LCRC and UCT on the Evaluation of Identi-Kits

Relevancy: Cohort x-y – Request for further information

To the hands of: Supt. XY

Dear Superintendent XY, dd.mm.yyyy

For our collaboration study on the evaluation of Identi-Kits, we need further information on the cases Capt. van der Westhuizen has provided us with since mmyyyy (Cohort x-y). As discussed before, the information needed can only be accessed via the LCRC database.

Please provide us with the following information on the cases listed below:

- Current legal status
- Fingerprint Evidence (checked for fingerprints?; fingerprints found?)
- Number of possible suspects under investigation

Thank you very much for your time and effort!

Heike Schmidt & Nils Maskow (UCT)

Police Station	Case Number	LCRC Number	Legal Status	Looked for fingerprints ?	Fingerprints found ?	Number of Suspects
Athlone	xxx/mm/yy	xxx/yy				
Atlantis	xxx/mm/yy	xxx/yy				
Atlantis CID	xxx/mm/yy	xxx/yy				
Atlantis	xxx/mm/yy	xxx/yy				
Bellville	xxx/mm/yy	xxx/yy				
Bellville CID	xxx/mm/yy	xxx/yy				
Bellville-South	xxx/mm/yy	xxx/yy				
Bishop Lavis	xxx/mm/yy	xxx/yy				
Bothasig CID	xxx/mm/yy	xxx/yy				
Bothasig	xxx/mm/yy	xxx/yy				
Camps Bay	xxx/mm/yy	xxx/yy				
Cape Town	xxx/mm/yy	xxx/yy				
Cape Town CID	xxx/mm/yy	xxx/yy				
Cape Town	xxx/mm/yy	xxx/yy				
Claremont	xxx/mm/yy	xxx/yy				
Delft	xxx/mm/yy	xxx/yy				

Appendix P-07

Answer sheet for the face matching

Please fill in your… Student No ______________

Gender: m ☐ f ☐

Age: ____

Race: black ☐ white ☐

Instructions:

In front of you is one envelope containing composites, and on top of it 8 photographs of white males. Spread out only the photos on the desk in front of you, so that you can see all of them well and at the same time. Look at each of the faces on the photos closely so that you remember how they look like. Then take the pile of face composites out of the envelope and choose, one after the other, which of the faces in the photos this displays, and put it in front of this photo. If you match several composites with one photo please pile them. You can match none, one, two, up till all composites to one photograph, it is up to you. But: you have to choose a photograph for each composite! **MATCH THE COMPOSITE TO THE PHOTOGRAPH THAT LOOKS MOST LIKE THE PERSON DEPICTED**. Try not to change your choice once you made it. **JUDGE EACH COMPOSITE INDIVIDUALLY!**

Once you're done with it take this sheet again and insert on the next page which composite (letter in the left bottom corner of each) you matched to which photo (number in the left bottom corner).

For each photo please indicate if you are familiar with the person on the photo. If so, indicate your grade of familiarity on the scale (1=barely familiar, 7=highly familiar).

When you are finished with this please indicate this to the experimenter.

Answers sheet – Which composite did you match to which photograph?

Photo No. 1
Matched composites: ____

Photo No. 2
Matched composites: ____

Photo No. 3
Matched composites: ____

Photo No. 4
Matched composites: ____

Photo No. 5
Matched composites: ____

Photo No. 6
Matched composites: ____

Photo No. 7
Matched composites: ____

Photo No. 8
Matched composites: ____

Do you know the person on the photo?

Photo No	YES	NO	IF YES: How familiar are you with him? (1=barely familiar, 7=highly familiar)
1			1—-2—-3—-4—-5—-6-----7
2			1—-2—-3—-4—-5—-6-----7
3			1—-2—-3—--4—-5—6-----7
4			1—-2—-3—-4—-5—-6-----7
5			1—-2—-3—-4—-5—-6-----7
6			1—-2—-3—-4—-5—-6-----7
7			1—-2—-3—-4—-5—-6-----7
8			1—-2—-3—-4—-5—-6-----7

Appendix P-08

Answer sheet for the face matching

Please fill in your… Student No
____

Gender: m ☐ f ☐

Age: ____

Race: black ☐ white ☐ coloured ☐

Instructions:

Projected onto the screen in front of you you'll see an array of photographs of white Males. Look closely at each face so that you remember what they look like. You might recognize some of the faces, please do not treat them any different than the unfamiliar faces in the photographs.

On the following pages you will see face composites. Each composite depicts one of the persons shown in the array of photographs. Please choose, one after the other, which of the faces in the photos the composite is depicting, and note this as your first choice. Write the red number in the bottom left corner of the photograph into the field "first choice". Subsequently do a second and third choice for each composite. After this, please turn the page and proceed with the next composite.

You may match multiple composites to a photograph, that is, you do not have to exclude a photograph from your choice once you have "used" it. But: You **must to make three choices for each composite**! Try not to change your choice once you have made it. **Judge each composite individually**

<u>Composite 1</u> – **Which photographs would you match to it?**

First Choice: __________

Second Choice: __________

Third Choice: __________

(please note: the remaining composites were presented in the same way as the first one displayed on this page)

Appendix P-09

Answer sheet for the face matching

Please fill in your… Student No _______________

Gender: m ☐ f ☐

Age: ____

Race: black ☐ white ☐ coloured ☐

Instructions:

Projected onto the screen in front of you you'll see an array of photographs of white Males. Look closely at each face so that you remember what they look like. You might recognize some of the faces, please do not treat them any different than the unfamiliar faces in the photographs.

On the following pages you will see face composites. Each composite depicts one of the persons shown in the array of photographs. Your task is to exclude the photographs that you are sure do not depict the composite. (Please decide which of the photographs might depict the person in the composite, and which might not do so.) **Cross out the number in the table related to the photographs that you want to exclude**. You can exclude/include as many photographs as you want to. After this, please turn the page and proceed with the next composite.

Try not to change your choice once you have made it. **Judge each composite individually!**

<u>Composite 1</u> – Cross out the numbers of the photographs you think do NOT depict the man in the composite below.

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56

(please note: the remaining composites were presented in the same way as the first one displayed on this page)

Appendix P-10

Collaboration study on composite quality

Participant Number: ____________ Version: __________

Inclusion criteria:

Are you familiar with a lot of people from American sports?
What sports?

Are you familiar with a lot of people from American show business?
What kind (i.e. movies, tv, singers, musicians, etc.)?

Demographic Data:

Gender: m f

Race: Black White Hispanic Asian Other: ________________

1. Composites

These are composites/reconstructions of people either from sports or showbiz in the US. We would like you to try and identify them, that is tell me who this composite depicts. You don't have to give the name of the person, any information that would identify only him is good (e.g. "played Forest Gump in this famous movie years ago", NOT "played a retarded guy in a movie once").

1	No	Yes, it is…
2	No	Yes, it is…
3	No	Yes, it is…
4	No	Yes, it is…
5	No	Yes, it is…
6	No	Yes, it is…
7	No	Yes, it is…
8	No	Yes, it is…
9	No	Yes, it is…
10	No	Yes, it is…
11	No	Yes, it is…
12	No	Yes, it is…
13	No	Yes, it is…
14	No	Yes, it is…
15	No	Yes, it is…
16	No	Yes, it is…

2. Pictures (photo naming)

These are photos of people either from sports or showbiz in the US upon which the composites were based. We would like you to try and identify them. If you identified them, please rate how familiar you are with the person? 1 is not really familiar, 7 is very familiar.

1	No	Yes, it is…	1-2-3-4-5-6-7
2	No	Yes, it is…	1-2-3-4-5-6-7
3	No	Yes, it is…	1-2-3-4-5-6-7
4	No	Yes, it is…	1-2-3-4-5-6-7
5	No	Yes, it is…	1-2-3-4-5-6-7
6	No	Yes, it is…	1-2-3-4-5-6-7
7	No	Yes, it is…	1-2-3-4-5-6-7
8	No	Yes, it is…	1-2-3-4-5-6-7

3. Composites (second naming)

We will now show you the same composites we showed to you in the first task. These composites were based upon the 8 celebrities you just saw photographs of. Please look at these composites again and try to identify them. Keep in mind that each composite is meant to represent 1 of the 8 celebrities you just saw.

1	No	Yes, it is…
2	No	Yes, it is…
3	No	Yes, it is…
4	No	Yes, it is…
5	No	Yes, it is…
6	No	Yes, it is…
7	No	Yes, it is…
8	No	Yes, it is…
9	No	Yes, it is…
10	No	Yes, it is…
11	No	Yes, it is…
12	No	Yes, it is…
13	No	Yes, it is…
14	No	Yes, it is…
15	No	Yes, it is…
16	No	Yes, it is…