UNDERSTANDING AND IMPROVING THE EFFECTIVENESS OF SKETCH FACIAL COMPOSITES

Heidi Kuivaniemi-Smith

PhD

University of Central Lancashire

Psychology

March 2023

Abstract

Constructing a facial composite sketch by a forensic artist is one of the oldest, yet least researched methods of facilitating an eyewitness's memory for an offender's face and depicting it based on their recall and instructions. While a lot is known about how faces are processed in relation to computerised composites systems such as E-FIT and newer holistic systems, little is known of the mechanisms of sketch composites and their effectiveness. There are two main ways of constructing a sketch composite; by using reference materials (e.g., pictures of facial features or hairstyles) as recognition aids to support recall or not using them and for the witness to rely on the developing sketch as a comparison point for their mental image of the face. How long an offender is seen is one of the key factors in the storage and retrieval of the memory of them and it is important to explore the role of reference materials (or not) when encoding duration has been short or longer. Experiments 1, 2 and 3 examined these two factors in a 2 x 2 study design. Overall, it was found that longer encoding (30 sec – 60 sec) led to more effective composites (resembled the target better/were more identifiable) than short encoding (5 sec). No overall benefit for the use of reference materials was found, however, it is indicated that they improve the composites after short encoding. Recall is a vital part in sketch composite construction, which can be enhanced in a composite interview with specific mnemonics and instructions. One of the most effective ones is mentally reinstating the context of the situation where encoding of the offender occurred by thinking about any cues from the environment, sounds, smells and one's emotions and moods. A more detailed version of this is to also verbally describe it to the interviewer, which has been found to be effective with computerised systems. It was hypothesised that sketches would also be improved. Two experiments manipulating context were conducted, which found some conflicting results. In one experiment, the detailed context reinstatement impaired the composite quality while another experiment found a benefit for sketch composites not using any reference materials. These results indicate that

sketch composites can be improved by providing witnesses with visual cues during construction and using memory enhancement techniques.

Table of contents

Glossary

CD-FIT – An early computerised feature composite system, now known as PRO-fit.

Cognitive Interview (CI) – An investigative interview that aims to elicit more accurate information without compromising the quality of recall.

Configural processing – Perceiving and recognising faces as a whole, based on the spatial arrangement of facial features rather than just individual features in isolation.

Conversation Management – A type of interview used in investigations.

Detailed CR – A mnemonic of the Cognitive Interview that guides a witness to focus on the context of the crime by verbalising it in detail.

E-FIT – A computerised feature composite system that uses colour images.

EFIT-V (EFIT 6) – A holistic composite system that uses colour images.

Enhanced Cognitive Interview (ECI) – The original Cognitive Interview with added guidelines (e.g., social dynamics).

Episodic memory – A form of long-term memory that captures the details of past events that one has personally experienced.

EvoFIT – A holistic composite system that uses greyscale images.

FACES – A computerised feature composite system.

Featural processing – analysing and recognising individual features rather than considering the face as a holistic or configural unit.

Fusiform gyrus – A region in the brain in the temporal lobe, that plays a crucial role in visual processing and recognition.

Gold standard – Guidelines set by Frowd et al. (2005b) to add more ecological value to facial composite research.

Holistic Cognitive Interview (H-CI) – A holistic protocol in the Cognitive Interview, that instructs a witness to reflect on the face's characteristics and make character judgments.

Holistic processing – Perceiving the entire face as a unified and meaningful structure, rather than focusing solely on individual facial features.

Identi-kit – An archaic mechanical composite system using line-drawing images.

Identikit 2000 – A sketch-like computerised feature composite system.

Inferior occipital gyrus – A region in the occipital lobe of the brain, that plays a role in visual processing and is involved in the recognition and analysis of visual stimuli.

Matt-lustre acetate – A transparent sheet to place over a mechanical composite image.

Mental reinstatement of context (MRC) – A mnemonic of the Cognitive Interview that guides a witness to focus on the context of the crime by thinking about it.

Minimal/usual CR – A mnemonic of the Cognitive Interview that guides a witness to focus on the context of the crime by thinking about it.

Mock witness – A participant in a study.

OFA "occipital face area" – A region in the occipital lobe of the brain, that is part of the broader network of brain regions responsible for visual processing.

Photofit – An archaic mechanical composite system using photographic images.

Physical CR – Recalling information in the same environment where encoding occurred.

PRO-fit – A computerised feature composite system that uses greyscale images.

Procedural memory – A form of long-term memory that enables people to learn and execute tasks.

Semantic memory – A form of long-term memory that comprises a person's knowledge about the world.

Superior temporal sulcus (pSTS) – A prominent groove in the temporal lobe, that is involved particularly in the processing of auditory and visual information, as well as social perception.

The Memorandum of Good Practice – Provides guidance to police officers and social workers responsible for undertaking video-recorded interviews with child victims or witnesses.

The Stepwise method – A method for interviewing children in cases of alleged sexual abuse.

Verbal overshadowing effect (VOE) – Verbally describing a face interferes with the recognition of it.

List of figures

Chapter 1:

Chapter 2:

Chapter 3:

List of tables

Chapter 1:

Chapter 2:

Chapter 3:

Acknowledgements

Firstly, I would like to thank my Director of Studies, Professor Charlie Frowd, for all the guidance in the area of facial composite research, which he has conducted so much over the past couple of decades. His expertise and passion for the topic, and friendly mannerism has helped me to go through this PhD process. Warm thanks to also my other supervisors Beth Richardson and Rob Nash, whose guidance has steered me particularly with my writing. Rob also helped me arrange data collection on composite construction for two of my experiments at the University of Surrey and Aston University, which I am very grateful for. Data was also collected at Reading University, and I thank Dr Rachel Pye for this opportunity. Thanks also to research assistants at Aston University and University of Winchester for helping me with data collection for composite evaluation such as composite naming. I want to extend my gratitude also to Reading FC and National Football Museum in Manchester for accommodating data collection for the composite naming tasks.

This was a long journey with a few interruptions to my studies for health and financial reasons. I am thankful that I was still able to continue with my studies after all the obstacles. I transferred from the University of Winchester to UCLan after about two years of studies, which meant that I have since then been mostly working on this thesis remotely, which added to the difficulties. Luckily most of my composite construction data had been collected at Winchester already as this is the most time-consuming part of the data collection and requires a lot of in person attendance. I appreciate the support of my former supervisors Manuela Thomae and Maria Uther

at Winchester. As can be seen, working on this thesis has involved a lot of networking too, which I enjoy.

Finally, having the support of my family and some good friends has also helped me enormously and without them I would not have been able to complete this huge undertaking.

RESEARCH STUDENT DECLARATION FORM

Type of Award	**Postgraduate Research Degree - PhD**
School	**School of Psychology and Computer Science**

1. Concurrent registration for two or more academic awards

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution.

2. Material submitted for another award

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work.

3. Collaboration

Where a candidate's research programme is part of a collaborative project, the thesis must indicate in addition clearly the candidate's individual contribution and the extent of the collaboration. Please state below:

4. Use of a Proof-reader

*****The following third-party proof-reading service was used for this thesis: *Anna Petkovic for Introduction Chapter (recommended by Prof Charlie Frowd)* in accordance with the Policy on Proof-reading for Research Degree Programmes and the Research Element of Professional Doctorate Programmes. In addition, Amanda de Grey proof-read the final thesis to detect any typographical and grammatical errors.

A copy of the confirmatory statement of acceptance from that service has been lodged with the Academic Registry.

Signature of Candidate HKS

Print name: Heidi Kuivaniemi-Smith

1 INTRODUCTION CHAPTER

1.1 Eyewitness memory and the legal system

1.1.1 Facial composite – a tool to aid identification of perpetrators in forensic investigations

Human sketching has not been a focus of intensive research. The main aim of this thesis is to improve sketch composites and improve their forensic utility by investigating the key elements of the process.

In an event of a crime, where no other evidence is immediately present, an eyewitness can play a key role in generating leads to the identity of the offender. Decades ago, mistaken eyewitness identifications have led to innocent people being convicted of a crime they did not commit, and this has forced the criminal justice system to review its practises concerning eyewitness memory. In the USA, the fallibility of eyewitness memory and the risk of suggestive identification procedures contributing to false identifications, was acknowledged in 1967 by the US Supreme Court (Clark, Moreland, & Rush, 2015*)*. In one case in England, a man called Laszlo Virag was identified by several eyewitnesses to have committed an armed robbery and was subsequently convicted (Devlin, 1976). He had a somewhat similar appearance to the actual criminal, Georges Payen, which made him vulnerable to misidentification. In another case, even when the witness spent around 30 minutes with the culprit, a false identification was made, which saw George Ince being convicted based on eyewitness evidence (Cole & Pringle, 1974). Due to these miscarriages of justice, a committee was formed and chaired by The Right Honourable Lord Devlin FBA in England, subsequently called the Devlin Committee

(Devlin, 1976), to commence a review of all aspects of law and procedures relating to evidence and identification in criminal cases (ACPO, 2009). The items in the report, that was produced following the review (Devlin, 1976), were however not made legally binding but rather remained as recommendations, leaving cases still vulnerable to misidentification (Davies, 1996; Davies & Griffiths, 2008). Similarly in Scotland, the Advocate's Guidelines have been used as a guide for facial identification procedures, but the fact that these are not statutory has repeatedly come up in appeals (Ferguson, 2017). Following the recommendations of a Bonomy Commmittee, Section 57 **-** Code of practice about investigative functions of the Criminal Justice (Scotland) Act 2016, now states that Lord Advocate must issue a code of practice on: (a) the questioning, and recording of questioning, of persons suspected of committing offences, and (b) the conduct of identification procedures involving such persons (Ferguson, 2017; see legislation.gov.uk for Criminal Justice (Scotland) Act 2016, Section 57).

Indeed, there are more recent examples in these miscarriages of justice, even when CCTV images have been available in addition to eyewitness identification, such as the case in Scotland in 2007 when a man was wrongfully convicted of a bank robbery (e.g., "The Man who police identified as a bank robber is freed from jail after DNA breakthrough"**,** 2009; Evidence-based Justice Lab, n.d) or when an innocent man was mistaken as a terrorist suspect and fatally shot by the Metropolitan Police in 2005 based on CCTV images (Lander & Bruce, 2018). An organisation called the Innocence Project, based in New York, is dedicated to fighting for justice for wrongfully convicted suspects, and many convictions have been exonerated due to DNA tests revealing mistaken identity (Innocence Project, 2023). In the UK, The Innocence Project London, is a member of the Innocence Project (US) and works towards the same goals (Innocence Project London, 2023). Despite the potential errors in eyewitness memory *(*e.g., Bruce et al., 1999; Burton et al., 1999*),* it is an invaluable source of information to an investigation. Eyewitness memory is now an enormous field of research, and concerns about potential errors have been addressed continuously to make the justice system fairer by minimising bias (e.g., Memon et al., 2011; Valentine & Heaton, 1999).

How does the above concern facial composites? Firstly, it highlights that memory of unfamiliar persons can be fallible, and with facial composites, there is an additional hurdle of retrieving a picture of the perpetrator's face from witness's memory through a sketch artist or composite system operator. This requires a careful procedure in facilitating witness memory. One of the observations that stemmed from reviewing the procedures concerning identification of suspects by the Devlin committee was contamination of memory, and thus it was recommended that photographs and facial composites (then Identi-kits, Photofits and artist sketches) should not be shown to witnesses prior to them viewing identification parades or confrontations (ACPO, 2009). These guidelines were included in the revised Code D, Code of Practice for the Identification of persons by police officers, which is part of the Police and Criminal Evidence Act 1984 (PACE) effective in England and Wales (ACPO, 2009; Home Office, 2012). Similarly, Code D in the Police and Criminal Evidence (Northern Ireland) Order 1989 is effective in Northern Ireland legislation (Department of Justice, 2015). Following these guidelines is important so that the attempt to tap into the eyewitness's memory of the face could be done with as little interference as possible. Naturally, how many faces a witness sees in real life during the interval of witnessing an incident and composite construction, is out of the control of the interviewer, but aiding to preserve the memory at the time of the interview is key to an appropriately completed process.

In Scotland, guidance concerning facial imaging is issued by Police Scotland as part of the Criminal Justice (Scotland) Act 2016. It differs from the England/Wales guidelines by, for example, recommending that a facial composite should not be constructed by a witness for at least 48 hours after the incident, to lessen the effects of any trauma (Police Scotland, 2021). It is mentioned that it is best practice not to show any photographs prior to composite construction to avoid contaminating memory. However, while it is acknowledged that in some situation this may be necessary, it is advised that this is likely to be scrutinised at court.

1.1.2 Facial composite as evidence

When the police first take a verbal description from a witness about what happened and find that no adequate video footage showing the offender's face exists (e.g., CCTV, mobile phone footage), and, in the absence of other identifying information, officers may request a facial composite, provided that the witness saw the offender's face clearly. This exercise requires an appropriately trained professional to interview the witness and construct a facial composite based on the witness's memory of the offender. For any case involving identification of a suspect, the police are required to follow the Turnbull guidelines that help to assess the quality of the identification (ACPO, 2009). These guidelines include questions such as "How long did the witness have the accused under observation? At what distance? Under what lighting?" (The Crown Prosecution Service, 2018). In the Scottish legislation, prior to requesting a facial composite, the officer in the case is similarly required to obtain a statement from the potential witness including the description of the suspect and particularly whether he or she saw the whole face of the suspect (Police Scotland, 2021). Officers should also assess the witness's ability to describe the suspect's face in detail. After the initial assessment by the officers, the sketch artist or composite system operator also needs to complete this assessment with the witness prior to proceeding with composite construction as part of the interview.

In terms of its evidential value, a completed facial composite image is a pictorial statement by the witness and any notes or sketches made during the interview are subject to disclosure under the Criminal Procedure and Investigations Act 1996 (CPIA) in England and Wales, and must be handled accordingly (ACPO, 2009). In the Criminal Justice Act 2003, effective in England and Wales, composites (such as sketch and Photofit or any other representation in a pictorial form) are considered hearsay evidence (see Criminal Justice Act (2003), section 115). With child witnesses, it is recommended that the interview (ABE- achieving best evidence) and composite construction, both by a computerised system or an artist's sketch, should be video-recorded (Ministry of Justice & NPCC, 2022). For a court to hear any new evidence in the same medium as the main evidence-in-chief (which is video-recorded) is likely to increase confidence in how the evidence was gathered, and importantly the video recording reduces the need for the child to repeat evidence in the witness box or via a live link (Ministry of Justice & NPCC, 2022). The final facial composite is circulated within the police and/or public for identification. For example, it could be shown in the BBC Crimewatch programme or posted in social media, although in the UK only around 10% of composites have historically been released to the media (Davies & Valentine, 2006). The aim of the composite is for someone to recognise the person depicted and for that lead to be investigated to see if other supporting evidence exists. Any one piece of evidence, including a facial composite, cannot on its own lead to a conviction of a suspect, and a case will not be accepted into court by the Crown Prosecution Service in the UK without sufficient and reliable evidence (Frowd, 2011). All parties involved in an investigation should have a good understanding that a composite represents a likeness of a suspect based on a witness's memory and is thus not a pictorial record created from the memory of a composite artist or a facial imaging officer, and is not comparable to a photograph (ACPO, 2009). According to ACPO (2009), a composite is "made up of various parts or blended" and likeness means "bearing close resemblance or having similar characteristics to the person portrayed". A facial composite can act as a key tool in providing a lead to an investigation that might otherwise have stalled.

Composites may only be considered if no other, more robust identifying evidence, comes to light. This means that most police officers do not regularly come across a situation where composites are appropriate, so it is likely that police forces are unclear about when and how to request a composite to be constructed. This may lead to weeks, or even months passing before a composite is requested, which is far from ideal since memory, particularly for detailed information, is likely to degrade during this time (e.g., Ellis et al., 1980; Gambell, 2006; Tulving & Craik, 2000). Also, if an event is not described soon after it occurred, over time there is an increased risk of recent events being confused with previous events (Quas et al., 2000). The situation could be improved by educating officers and other police staff about facial composites.

Facial composite research aims to improve the likeness of the composite to the subject being depicted based on eyewitness's memory, ultimately increasing the utility of composites in forensic investigations. Any improvements based on evidence from research can then be applied to the procedures of constructing composites with witnesses. On the other hand, if something is found to affect the quality of the composite (likeness) negatively, procedures can be amended accordingly. An artist’s belief to their technique of constructing a composite is required to be neutral in research and not biased, as it often is in the field. When a practising sketch artist uses a technique that he or she has honed over the years and this has led to successful composites (anecdotal evidence of success rates), it is natural for bias to be present. This, however, may mean that potential improvements to the technique may not be considered.

How the sketch composites have been used in the past for investigations and how this may have affected their use in research is explored in the next section, along with other composite systems.

1.2 The history of facial composites

1.2.1 Sketch composites

The manual method of drawing using paper and pencils (or other drawing utensils) has been utilised in the forensic field for over a century (Taylor, 2001). Drawings of suspects were published, for example, in the 'wanted' posters in the USA from the 1800s (Taylor, 2001). In 1911, an artist impression that depicted the current appearance of a man wanted for the murder of his wife was circulated by the Metropolitan Police and as a result a passenger on a transatlantic liner identified the man depicted (Davies & Valentine, 2006). Sketching was also the first method used for facial- composite construction when technology for it did not yet exist. As early as 1881, an artist created a caricature style sketch of murder suspect Percy Lefroy

Mapleton, which led to his arrest (Wood, 2019). Mapleton was seen covered in blood on a train in England after having murdered his fellow passenger Isaac Gold and claimed to have been attacked by another passenger. He became the prime suspect after the body of Mr. Gold was found soon after the murder. The artist's impression, which was created with the help of a person who knew the suspect well, was published in a newspaper alongside the police officer's description, which resulted in information being given of the suspect, who was subsequently arrested. Another sketch that was drawn from memory and descriptions of a witness, who described a previously unknown person to them, led to the identification and arrest of the Wall Street bomber in 1920 (Taylor, 2001).

There have been autobiographies (e.g., Boylan, 2000), novels (e.g., Stuart Parks, 2017) and practical manuals on different methods for sketching including digital sketching (e.g., Gibson, 2010; Mancusi, 2010; Murry, 2018; Taylor, 2001) written by experienced and accomplished sketch artists. However, no consistent empirical evidence on the sketching methods exists, which is rather surprising given that considerable anecdotal evidence indicates that the sketch composites can help to identify perpetrators and have done so in many high-profile cases and serious crimes (e.g., Boylan, 2000; Gibson, 2010; Taylor, 2001). One reason for the lack of research is likely to be that sketching methods vary widely and are adapted by each artist to their own desired way of working, thus making methods difficult to measure and compare. The composite construction interview is also an intuitive process during which the artist will interpret not only the verbal information given by a witness but also his or her body language and any gestures made that help to support the descriptions (e.g., Taylor, 2000). Body language is an important part of our interaction and can certainly help as part of gaining mutual understanding. Since every witness and interview is unique, it is impossible for a sketch artist to follow a strict structure and plan everything prior to the interview.

Davies's (1986) history of sketch artists in North America and Britain mentions that while there was a growing number of police artists in Canada and USA, they were still very much a minority as most composites were constructed using mechanical systems such as Photofit and Identi-Kit. According to MacDonald (1984), there were about two hundred full-time artists for all 2000 law enforcement agencies in the USA, and a survey of Canada's police showed that only 17 of the 133 personnel involved in identification duties were sketch artists (Arnold, 1982). These numbers appear to have been getting continuously lower and Davies and Valentine (2006) state that the field of sketch composites is a "dying art". More recently, forensic artist Catyana Falsetti (C.Falsetti, personal communication, April 22, 2020) estimates that the number of full-time forensic artists in the USA reduced to about 49 in 2014, the highest number being at the FBI, the National Center for Missing and Exploited Children, and the Los Angeles Sheriff's Department. Many other departments have one to three artists according to Falsetti.

In the 1980s, the field of sketch composites started to become more organised, and in 1983 the first training manual, The Law Enforcement Composite Sketch Artist by George Homa (a resident sketch artist with the New Jersey State Police), was published (Davies, 1986). This was followed by the first international convention of police artists in New York in 1984 (Davies, 1986). Artists have been employed as required by the police in the UK, but only one force, Lancashire Constabulary, had a full-time sketch artist working directly with witnesses in the North American way (Davies, 1986). Sir Richard Jackson (1967), an assistant Commissioner in Scotland Yard and a President of Interpol, was introduced to one of the mechanical systems, Identi-Kit, in 1959 and following its successful trial use in serious crimes in 1961, Identi-Kit was leased to those police forces in the UK who wished to use it. The Interpol General Assembly however did not recommend its use internationally as the kit was only applicable to Caucasian faces. Jackson remarked that the general public in Britain was, in his opinion, irrationally apprehensive about the use of this new tool due to fear of their privacy being violated or composites leading to wrongful convictions. He emphasised that an Identi-kit image alone could not lead to conviction but would need an identification line-up to be assembled if the composite was recognised. Despite this, one unfortunate case exists in which an innocent man resembling a composite was convicted in the Netherlands in 1975 and spent a year in prison before the injustice was corrected (Shepherd & Ellis, 1996). Jackson maintains that constructing pictures from eyewitnesses' descriptions was nothing new at the time. He mentions one example where he was impressed by a good likeness of a sketch drawn in profile view of a suspect that showed the man's hooked nose with glasses. This appears to have been a rare case as he then states that in Britain, unlike some other countries, it was not normal practice for artists to work directly with witnesses but from descriptions handed to them. An existing Photofit composite could also be turned into a sketch by an artist (Davies, 1986).

Jackson (1967) compares this indirect method to the process used with Identi-Kit and claims that a witness is too inclined to accept the completed sketch that is shown to him or her, while they can work on the Identi-Kit composite interactively until he/she is satisfied with the likeness. This appears to have been an anecdotal statement rather than observation based on any empirical evidence, which was lacking at the time. However, it makes sense since the witnesses might have believed that they had already done their part with the descriptions, and the process was out of their hands. This is entirely opposite to how an artist interacts with the witness now, following the principles of the cognitive interview, a type of investigative interview that aims to elicit more accurate information without compromising the quality of recall (Fisher & Geiselman, 1992). The witness should be made to feel that they are fully in charge of how the sketch turns out and that the artist is helping them to achieve that objective, as discussed later. Melissa Little, a British sketch artist, remembers that when Photofit was used in the UK, sketch artists were employed to add alterations to a composite if the kit did not contain a facial feature that was sufficiently similar to that required by a witness (M. Little, personal communication, June 24, 2021).

Current situation of sketch artists

There are no international standards for sketching in terms of craftsmanship (Davies & Valentine, 2006), and, as such, it is the sketch artist's own decision about which media should be used in each case. Composite sketches, which are distributed publicly, appear to be created mostly in grey scale by pencil or charcoal, but colour media is used sometimes. Digital sketching (See Figure 1.1), often involving Adobe Photoshop or other paint packages, is becoming popular among sketch artists due to its apparent advantages in image editing over the manual technique (e.g., Murry, 2018). While there are probably as many sketching techniques as there are sketch artists, since through their own experience the artist shapes their technique into a unique approach that suits them, there appears to be two main types of sketching methods: one that uses reference materials such as facial images/mugshots or facial features in catalogues, and one that relies on a witness's description and the evolving sketch that the witness sees. There may be some fundamental differences between these two techniques affecting the outcome of the composites, and thus it is one of the key topics to explore in this thesis. This motivation is also with an aim to recommend some guidelines for the sketch composites.

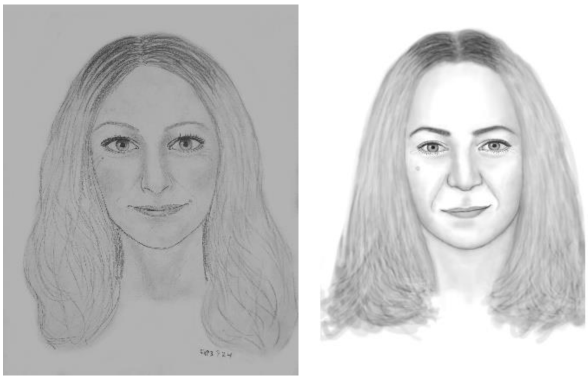


Figure 1.1 On the left, a traditional sketch drawn on paper with pencils, and on the right, a digital free hand drawing created on an iPad using a Procreate art program. Author's own images.

1.2.2 Mechanical composite systems

Since a skilled forensic sketch artist has always been a rarity, solutions have been sought for more people to construct composites, and to modernise the technique and reduce costs. The first attempt to make the composite construction process less dependent on the level of artistic skill of the person carrying out the interview was the creation of mechanical systems. These systems required manual assembling of pictures of facial features. The most well-known systems that have also been extensively tested in laboratory settings are Identi-Kit and Photofit.

Identi-Kit was developed by detective Hugh McDonald in the Los Angeles Identification Bureau in 1959 (Figure 1.2). It consisted of 500 features of chins, mouths, eyes, noses and hair (Identi-Kit Solutions, n.d.), presented as line drawings on transparent celluloid sheets (Laughery et al., 1977).

Figure 1.2 The original Identi-Kit. (WorthPoint, n.d.)

Shepherd and Ellis (1996) describe Identi-Kit II, an updated version of the kit published in 1975, that contains photographic elements like Photofit. It contains female features and transparencies for varying tonal values for non-white subjects. To improve the available examples of hairstyles, multiple hairstyle transparencies can be stacked on top of each other to use different sections from more features than one. The guidance in this version states that, if amendments to facial features are made, the operator should do this without a witness looking and only show the composite to them once the feature is in place so that the witness sees the face as a whole (Shepherd & Ellis, 1996), a procedure that has been shown to be more effective than when witnesses look at facial features out of context of a complete face (e.g., Skelton et al., 2015).

Photofit (Figure 1.3) was developed in 1969 in England (Laughery et al., 1977) by Jacques Penny (Davies & Christie, 1982) and has always contained photographic features, hence its name. Facial features were placed together on a specially constructed board to create a face (Laughery et al., 1977).

Figure 1.3 The original Photofit. (Greater Manchester Police, 2009)

An experiment conducted at Aberdeen University using Photofit discovered that several female features were consistently confused with young male features (Gibling & Bennett, 1994). Subsequently, a number of female Photo-fit features were added to the male selection as a supplement, termed the *Aberdeen Supplement*, and these additional features were claimed to help witnesses greatly in constructing adolescent male faces (Gibling & Bennett, 1994).

There are other early systems such as Facial Identification System (FIS), which allows witnesses to work on the composite without a technician, and the Minolta Montage Synthesizer, which was developed and widely used in Japan but only used in a couple of trials in the US (Laughery et al., 1977). FIS appears to be the first mechanical system that enables witnesses to select facial features in the context of a whole face (Luu & Geiselman, 1993). The kit consists of a stacked series of artists' illustrations in four rows, top portion (hair), second (eyes) and third (noses) portions and the bottom portion (jaws and mouth). The basic version also contained transparent overlays to alter the original composites (see, Luu & Geiselman, 1993, for a detailed description). Luu and Geiselman (1993) state that only limited research exists on FIS but hypothesised that the more holistic aspect should improve the effectiveness of composites compared to the other mechanical systems. Their study had two independent factors: order of construction (seeing features together in the whole face context vs. seeing features individually) and memory retrieval procedure (Cognitive interview (CI) vs. standard interview 'think about the face of the suspect'). Participants saw a short film of a robbery and constructed a composite immediately afterwards. Two different approaches of constructing the composites were used, 1) holistic-to-specific, which required the participant to construct the face, FIS-style, by selecting facial features in a whole-face context, and 2) specific-to-holistic, when facial features were selected in isolation of each other until all features making a complete face were selected. Two independent judges, who were shown examples of a good and a poor composite of an unrelated target as a reference, rated all 96 composites. As hypothesised, it was found that the CI enhanced composite construction when combined with the holistic-to-specific method but had a negative (but non-significant) effect when the specific-to-holistic approach was used. Authors concluded that the CI is a promising retrieval aid for composite construction when the procedure used to create the face is holistic in nature (i.e., when features are viewed in a whole-face context). One limitation of the study was that only two people were asked to evaluate the composites, potentially giving a result that may not be as generalisable as if a larger group were recruited to carry out the assessment (as is the usual case with assessment for face construction: e.g., Frowd, Bruce, Smith & Hancock, 2008). Most of the literature appears to focus on Identi-kit and Photofit and, while they are clearly outdated now as composite systems, their evaluation contributes robustly to the large body of research on facial composite systems.

1.2.3 Computerised composite systems

An early attempt to use computer graphics to construct a facial image involved a software system called WHATSISFACE in the 1970s. It contained prestored facial features that could be edited, and it combined aspects of computer graphics, artificial intelligence and pattern recognition to construct a line drawing type image (Gillenson & Chandrasekaren, 1975). However, the authors pointed out that the intention was not to produce images from a witness's memory with this system but from a photograph in front of the user. A digitised prototype composite system using Photofit features was developed by the Computer-Aided Design Centre (CADC) in Cambridge (Davies & Valentine, 2006). Contrary to expectation, this system did not perform any better than the manual Photofit (Christie et al. 1981). As computer technology continued to improve, there was a natural shift to developing composite systems that would be more versatile and had larger databases of facial features compared to the mechanical systems. This situation remained until 1994 when Electronic Graphics (more recently Aspley Ltd.) introduced the Electronic Facial Identification Technique (E-FIT), which was effectively an electronic version of Photofit developed by the Home Office and Aberdeen University (Gibling & Bennett, 1994). Until then, Photofit, despite its shortcomings, was still the most widely used composite system; in England and Wales 38 of 42 forces have used Photofit since its launch in 1970 (Gibling & Bennett, 1994). E-FIT is now probably the most well-known system in the UK, also familiar to the general public; composites are still generically being referred to as E-FITs even if the composite had been constructed with another system. E-FIT is based on its predecessor, CADC (Davies & Valentine, 2006) and uses photographic colour images (see figure 1.4). PRO-fit, originally known as CD-Fit (Bruce et al. 2002), is a similar system. Composites are usually produced in greyscale (see figure 1.5) and E-FIT is an exception in this arena rather than the norm. Although colour has been shown to affect recognition memory, the same does not appear to apply to identification of faces (Kemp et al., 1996). Kemp et al. (1996) acknowledge however that in the recognition task, using the same picture in the learning and testing stages could have played a part in enhancing recognition rates by encoding pictorial cues (same items in the pictures) rather than the recognition of the face itself, (which requires a different picture of the face). For facial composites, it has been demonstrated that a face presented in colour does not promote an overall more identifiable image than when presented in greyscale (e.g., Frowd et al., 2006).

The early computerised composite systems, such as Mac-A-Mug, are referred to as feature-based systems (e.g., Davies, Van der Willik & Morrison, 2000). However, compared to the mechanical techniques, systems such as E-FIT are a step towards face construction allowing a natural holistic face processing to greater extent since it is possible to select individual features in the context of the whole face (much the same as FIS), which has long been considered beneficial in the use of these systems (Davies et al., 2000; Davies & Milne, 1982). More recent research supports this practice. Skelton, Frowd and Speers (2015) compared the method of selecting isolated features and features in the whole face context using the PRO-fit feature system. In Experiment 1, two groups of forensic psychology students (twenty participants in total) participated in the study during a seminar on facial composites. They were briefed about using PRO-fit composite system according to their allocated condition, which was either construction of a composite whilst all facial features were visible, or in isolation from each other. Both groups viewed a still image of a celebrity for 1 minute (10 celebrity targets were used, each participant constructing a different face per group), after which they were instructed to write down what they could recall of the target face on a description sheet containing prompts for all features (e.g., facial shape, nose, mouth). They also were given access to written instructions on constructing the composite. The whole-face composites were constructed as described in Fodarella et al. (2015), and in the isolated feature condition, all features were selected separately, after which the whole face was revealed. The selected features could then be resized or repositioned. A different set of participants (N = 18) evaluated the composites by attempting to name them. Due to the composites having been self-constructed by other participants and not by a trained composite operator, they were expected to be of worse quality and thus, naming participants were given a name list of all targets and they attempted to name the composites from both conditions with the help of the list. As hypothesised, composites constructed in the whole-face context were named significantly better than those produced using isolated-feature selection. Since Experiment 1 was lacking in ecological validity, Experiment 2 was conducted to mimic a real crime scenario as all facial composite research following the gold standard by Frowd et al. (2005b). The targets (ten UK footballers) were unfamiliar to the composite constructors and composites were constructed approximately 24 hours after viewing the target. PRO-fit was now operated by the experimenter instead of the participants themselves (N = 20), who were interviewed using a Cognitive Interview, as in real cases. The composites were evaluated by two different naming tasks by new participants (N = 36). Firstly, participants attempted to name the targets (UK footballers) by seeing the whole face of the composites. Twenty-four participants saw all composites from both conditions in a within-participants design. In a second task, twelve participants attempted to name composites by only seeing their internal features to determine whether the internal features were more accurately constructed in the whole face condition than in the isolated-feature condition. In this task, a list containing the targets' names was given to the participants, as was the case in Experiment 1 naming task, due to the composites being predicted to be more difficult to name with removed external features. It was found that the correct naming rate was significantly higher for composites constructed in the whole face context than for those constructed while seeing the features separately. The same benefit of the whole face context was found in naming of the internal-feature-only composites. The possibility of constructing a composite in a whole face context appears to have been a positive step towards achieving composites with a better resemblance to the perpetrator. This aspect could offer some benefit to sketching also, which would seem to be worth investigating. (This suggestion will be linked to the question mentioned earlier of whether to use or not use reference materials as recognition aids for the witness constructing the composite). In the United States, systems that rely on featural processing (i.e., selection without a whole face context) have remained popular despite empirical evidence pointing to these systems performing worse than systems with a more holistic aspect (e.g., Frowd et al., 2007 2008). For example, the featural system FACES tends to produce less effective composites than PRO-fit (Frowd et al., 2007), the latter of which uses selection in the context of a whole face. Identi-Kit 2000 and FACES 4.0 are among the most popular systems currently used by law enforcement in the United States (Frowd, Erickson et al., 2015; McQuiston-Surrett, Topp & Malpass, 2006).

Figure 1.4 E-FIT composite.
(ACPO, 2009)

Figure 1.5 PRO-fit composite.
Author's own image.

Current face construction practises on computerised feature systems

Seeing too many reference images, whether in terms of whole faces or isolated facial features, is likely to be confusing to a witness. Face recognition research supports this idea. Bruce and Young (1986) found that recognising a face from a long list of mugshots declines dramatically at later positions when target faces are presented. The authors claim that this is due to the interference to memory by seeing many faces. When mugshots were limited to a smaller sample, witnesses were able to recognise faces better compared to when they viewed the entire photo album (Shepherd, 1986). To limit interference, as part of best practice (e.g., Fodarella et al., 2015), a composite practitioner narrows down the number of available individual facial features to show the witness only those that are based on the witness's initial description. Otherwise, fatigue is likely to occur. After this, the practitioner explains to the witness that the initial face might not look like the offender yet and that the idea is for the witness to look through the features and select ones that looked most like the offender's features. The witness continues selecting features (in context of a complete face) until he or she is satisfied that the best likeness has been reached.

Additional artwork (e.g., refinement to the shape and appearance of facial features) can be added to the composite image by the composite practitioner under the direction of the witness, as required.

Bruce et al. (2002) compared the effectiveness of combining or “morphing” PRO-fit composites constructed by different participants to composites constructed by one individual. In Experiment 1, four participants constructed composites of a familiar celebrity and an unfamiliar target, while viewing the target face and from memory. When the target face was in view, the participants described this to the experimenter who constructed the composite with a time limit of 30 minutes. In the memory condition, participants constructed another composite with the experimenter after seeing the target image for 30 seconds. Morphing (averaging) was first applied to two composites, making the composite 50 % of each face (2-morph). Another two composites of the same target were generated in the same way. These two morphed composites were then morphed together, containing 25% of each face (4-morph). In addition, weighted composites were generated which depended on the confidence ratings the composite constructors gave when the composites had been completed. The percentage of morphing was applied to each composite accordingly. In the experiment, there were four individual composites, four 2-Morphs (two weighted, two unweighted) and two 4-Morphs (one weighted, one unweighted), thus, ten composites in total for each target in each condition.

Composites were evaluated for likeness to targets by forty new participants on a scale of 1-10. Separate booklets containing familiar or unfamiliar faces were made and each composite was presented alongside its target face. Familiar and unfamiliar faces were evaluated separately. The composite constructors' own ratings of the composites did not influence the likeness ratings in general. The individual composites were rated the lowest. The 2-Morph composites were rated better, and the 4–Morph composites received the highest ratings on average. By item analysis revealed that 4-Morphs were still better than both 2-Morphs and the highest rated individual composites. Composites of the familiar faces were rated significantly lower than composites of the unfamiliar faces. This supports a likeness rating procedure, which is carried out by participants who are unfamiliar with the target pool, as otherwise the ratings are less likely to be a true reflection of the level of perceived likeness of an unfamiliar face. With familiar faces, there are likely to be higher expectations especially on the accuracy of the internal facial features, which may be an unfair evaluation for faces only seen for a short duration prior to composite construction. The composites were also evaluated in identification line-ups. Participants saw line-ups of six faces with similar hairstyle and of approximately same age for each target and attempted to identify the target prompted by the best composite, the worst composite (based on the likeness rating results), the best performing morph-composite or all four individual composites. Separate line-ups were created for the familiar and unfamiliar faces. Morphed composites performed best for familiar faces, however, due to infrequent correct target selection for unfamiliar faces in general, there were no differences found in this group.

In Experiment 2, a more ecologically valid design was used. The composites were constructed by participants (N = 16) unfamiliar with the target they encoded. The composites were constructed in both frontal view and 3/4 view. Unweighted (not rated by the constructors) composites were morphed for all four targets. The composites were rated for likeness by new participants (N = 20) who did not know the targets. The four individual composites and the morphed composite of these were presented to the participants in the same way as in Experiment 1. The morphed composites were rated higher than the individual composites, significantly so for two of the targets. Best individual composites were identified, and the morphed composites were found to be as good as these and significantly better than the average individual composite.

Thirty-two participants familiar with the targets also tried to identify the targets from a set containing the best, the worst, morphed or all individual composites. No significant differences were found between the composite types; however, there was a similar trend as in Experiment 1 and the morphed composites performed similarly to the best composites. In addition, though, seeing all composites performed best. Finally, sixty-four participants attempted to identify the targets unfamiliar to them from a line-up consisting of distractors that were selected based on the composite constructors' verbal descriptions. Composites were evaluated in four groups again: the best, the worst and all individual composites, and the morphed composites. No main effect was found for the composite types, although there was a trend in the morphed composites performing the best once again. Interestingly, the best and worst composites were identified at a similar rate. When data from all other conditions were combined, the morphed composites were identified significantly more frequently than the rest. As a result of positive findings for the morphing manipulation the ACPO face identification guidelines include instructions for morphing together two or more composites produced by different witnesses of the same unknown suspect (ACPO, 2009). The guidelines also state that it may not be possible to morph some artist sketches with computerised composites (*ibid*).

1.2.4 Next generation evolutionary systems

The knowledge that adults tend to recognise upright, intact faces in a holistic manner (e.g., Tanaka & Farah, 1993; Tanaka & Sengco, 1997) has resulted in the development of holistic composite systems (e.g., Frowd, Hancock, & Carson, 2004; Gibson et al., 2009; Tredoux et al., 2006). While featural processing relies on attending to individual features (Farah, 1991) and tends to be activated with unfamiliar stimuli (Rakover, 2002; Rhodes, Hayward, & Winkler, 2005), configural processing relies on attending to spatial relationships between the features (Tanaka & Sengco, 1997) and is activated by familiar stimuli, which faces in general are (Farah, Tanaka, & Drain, 1995; Rakover, 2002; Rhodes et al., 2005; Wilford & Wells, 2010). The distinction between holistic and configural processing will be explained in detail in the following chapter; however, in brief for clarity here, seeing a whole face as opposed to individual features separately is a principle upon which holistic composite systems are based. These systems aim to make face construction more effective for the witness by showing screens of whole faces, or whole face regions, to select (e.g., Frowd, 2021). The process then continues by the algorithm “breeding” the selected faces together and the image evolves gradually towards a better likeness to the mental image of the witness, as new screens of faces are presented. Holistic recognition of the face tends to withstand the degradation of memory better than face recall as the time from encoding to retrieval of the mental image of the person increases (Davies, 1983), and the potential of these systems to perform better than feature systems therefore lies in the fact that they tap into the recognition memory rather than recall. When there has been a long delay between the witnessed event and the construction of the composite, it is likely that holistic systems work better.

Hancock (2000) presented a prototype of the holistic system EvoFIT, which uses a principal components analysis (PCA), a statistical method for simplifying large datasets that had already been extensively used for face recognition and analysis at the time (Frowd, Hancock & Carson, 2004). Frowd et al. (2004) give a good overview of the technical aspects of the early EvoFIT. Early applications of the PCA to faces produced blurry images because the alignment of facial features within the facial model was simplified and the features were not always in the same place (Sirovich & Kirby, 1987). Co-ordinate points located around the major facial features (eyes, eyebrows, nose, and mouth) and the outline of the head, including the ears, chin, and jaw by Craw and Cameron (1991) enabled a triangular mesh to be formed from the averages of each point, which in turn allows all the positions of the features within the image to match by keeping the equivalent triangle in each face the same shape. The resulting images were referred to as “shape-free”. Hancock et al. (1996) realised that the control point information could be used as a part of a PCA that models the relational aspects of the face (e.g., the distances between facial features), which would lead to forming the shape of the face. A second model, referred to as the texture model, is what information remains in the image after it is made shape free. The texture part makes the images clearer. An evolutionary algorithm then produces a new set of faces from the selected parameter sets by recombination and mutation, resulting in faces that combine aspects of the previously chosen faces. As this process continues with new faces that are selected, the face evolves towards a desired likeness of a subject.

EvoFIT (figure 1.6) has been extensively researched since its development into a practical composite software (Frowd, 2021; Frowd, Hancock, & Carson, 2004; Hancock & Frowd, 2002). Similar to other facial composite systems, the performance of EvoFIT left a lot of room for improvement in its early days (Frowd et al., 2004; Frowd et al. 2005a, 2005b), because despite training the operator, the software was complex to use (Frowd et al., 2005a). This was found to be the case especially for the Feature Shift tool that enabled the shape and placement of features to be manipulated, an aspect of the system which has since been improved (Frowd et al., 2005a). The other key milestones in improving the user friendliness and achieving a better likeness of the composite to the subject include focusing of internal features during face construction (e.g., Frowd et al., 2012), using holistic tools after a face has been evolved (e.g., Frowd et al., 2006; Frowd et al., 2010), and focusing on the eye region at the initial interview (e.g., Skelton et al., 2019) and face construction (e.g., Fodarella et al., 2017).

Figure 1.6 An example EvoFIT composite. (EvoFIT, n.d.)

Other holistic systems include EFIT-V (rebranded as EFIT-6; (figure 1.7) and ID, but they have limited supporting research to assess their performance using a realistic methodology that includes a long delay (at least 1 day) from crime to composite construction (Frowd, et al., 2019). EFIT-V was initially called EigenFIT (Gibson, Solomon & Pallares-Bejarano, 2003). A brief explanation of the technical aspects can be found in Solomon et al. (2004), but in summary it is a similar system to EvoFIT, which uses an interactive genetic algorithm. It starts by a random generation of faces, to which the face constructor assigns fitness scores in order to indicate how similar the faces are compared to the target face. The algorithm then selects suitable faces for further breeding by employing “mutation” and “crossover” genetic operators and removes faces of lower fit, replacing them with faces of better fitness score. This follows the first recorded GA model used for facial composite generation by Caldwell and Johnston (1991). EFIT-V achieved a commercial status in 2007 (Solomon et al., 2012). It employs two core elements: the construction of a statistical appearance model of human faces and a stochastic search algorithm. In a trial funded by the UK Home Office that took place in two UK police forces, EFIT-V successfully provided intelligence, arrest or detection in 30% of cases in which it was used. In these cases, witnesses had been judged incapable of producing a composite using a standard, feature-based system used in the police at the time (George et al., 2008).

Figure 1.7 An example E-FIT6 composite. (Tsourrai & Davis, 2020)

Current face construction procedure on EvoFIT

Considerable research has assessed and shaped the EvoFIT system (e.g., Frowd et al., 2006, 2012; Frowd, 2021), and parts of the procedure are unique among holistic systems. Internal features such as eyes, nose and mouth are constructed first, with external features such as hair and ears masked and then selected towards the end of the process, a procedure that has been found to lead to more identifiable composites (e.g., Frowd et al., 2012)—as explained in more detail later in this chapter. Also, a witness is instructed to focus on the upper half of the face as this region has been found to further improve correct identification of the composites (Fodarella et al., 2021). Further, witnesses should not base selections on face width, but ignore this aspect of the face, as it can be altered with tools later on in the construction process (Fodarella et al., 2015). Again, for a detailed composite construction process, see Fodarella et al. (2015).

While all these composite construction methods differ from each other, they all have the same goal; to produce an image from the witness's memory that resembles the perpetrator as much as possible, so that ultimately someone within the police or from the public would recognise them. How the face has been processed, how it is stored in memory and how it is later retrieved from memory, are the main aspects enabling composite construction. An overview of these stages is given in the following sections. While the first two aspects occur automatically, retrieval of the memory as a picture of a face is a conscious effort and relies on different principles of face processing, depending on which system is used, some of which have a more of a focus on features while others have a more global focus. Due to different methods of constructing a sketch is used by sketch artists (using reference materials as recognition aids with a witness or not), it is important to start investigating how these methods match the encoding of the face, and thus support the retrieval of the memory of it.

1.3 Memory for faces

When a facial composite is required, it is often based on one specific event during which the perpetrator was seen. As discussed above, encountering an unfamiliar person in a crime situation can easily lead to misidentification due to the fallibility of eyewitness memory. When there is no known suspect, there is no face to identify yet, and the case depends on generating a lead to the identity of the perpetrator. Memory of an unfamiliar face typically contains some error; thus, a facial composite is not a portrait, but aims to achieve as much resemblance to the perpetrator as possible. Detailed information is required to be elicited from the witnesses' memory, which is demanding and difficult for him or her, not least since detailed information is vulnerable to degrading faster than overall information (e.g., Ellis et al., 1980; Gambell, 2006; Tulving & Craik, 2000). The first hurdle for achieving an effective composite is that a face may have been seen for a short period of time and may be restricted to one or two viewing angles, as opposed to a familiar face that has been exposed to in a 3-dimensional space on multiple occasions. The following discusses the differences between familiar and unfamiliar face recognition for a deeper understanding of the hurdles confronting constructing of a facial composite of an unfamiliar face.

1.3.1 Encoding an unfamiliar face

Despite the advances in forensic science in the last decades, it has been estimated that physical evidence linking an offender to a crime scene is available approximately only in 2% of the cases (Peterson et al., 2010), and, as such, eyewitness evidence plays a crucial role in investigations (Pike et al., 2019). A criminal is usually only seen for a short period of time, and thus the opportunity to encode this individual's face is limited. It is possible, though, to still construct a facial composite from witness's memory even after seeing the face for a short time, and there are examples in real cases that this can be achieved successfully (e.g., Boylan, 2000). In case studies, many variables (esp. estimator variables) cannot be controlled (e.g., length or severity of crime) and while it can be clear from an event that encoding may have occurred for a short duration, the exact time is not usually known. Laboratory studies, on the other hand, can manipulate or otherwise control this variable to better understand the impact of short and longer encoding durations.

Visual perception is the main basis for creating a memory of a face, and the processing of the face is likely to be influenced in terms of configural information (distances between features) when encoding duration is short, as there is limited time for processing individual features (e.g., Goffaux & Rossion, 2006). How this variable affects composite construction is an important interest of this chapter in the context of use (or not) of reference materials. As Wells and Olson (2003) suggest, the attention paid to a face is probably more important than exposure time per se. This is especially the case when a weapon is present, as that object can take the focus of attention away from the perpetrator's face (e.g., Erickson et al., 2022; Loftus et al., 1987; Steblay, 1992; Tooley et al., 1987), such that it might not be encoded satisfactorily for a later retrieval of the face to be successful (or to provide a less effective outcome). The author recalls an event that occurred in 2004, where two robbers held her and her companion at a gunpoint and knife in a post office robbery. It is possible that alongside other factors, the presence of the weapons led to the author and the other person to have different memories of the colours of the helmets on one of the robbers. However, other circumstances, such as the author having had a brief verbal exchange with one robber before the robbers escaped, enabling the author to look at this robber directly in the eyes, could have played a role. The witnesses are not aware of the robbers being arrested after the event; thus, the accuracy of the witness memory remains unknown. In this case only the eyes of the robbers were visible, increasing their chances of remaining unidentified. Interestingly, though, the author still has a mental image of the robber's eyes, possibly helped by the fact that he expressed some remorse to his actions, which was evident in the expression in his eyes.

Challenges described above do not apply to familiar faces, which can be recognised with ease, and generally in less than half a second (Bruce & Young, 1986). In most cases of composite construction, the perpetrator is unknown to a witness and recalling an unfamiliar face is notably more difficult than a familiar face, and prone to errors (e.g., Bruce et al., 1999; Burton et al., 1999). There is a distinction in the way the brain processes these two types of faces, and their neural substrate resides in different areas of the brain (e.g., Johnston & Edmonds, 2009). Recognising unfamiliar faces is made yet more difficult by changes in pose, expression, and context (e.g., Bruce, 1982; Hancock, 2012; Johnston & Edmonds, 2009) since these aspects have not been learned to the same extent as for familiar faces, often leaving reliance on one static view of the face (Hole & Bourne, 2010).

Although some exposures to a face can make it become more familiar, truly robust representations are developed only after many thousands of encounters with the face, especially under different conditions (Hole & Bourne, 2010).

The three-stage model of the recognition process (Bruce & Young, 1998) includes: i) matching a record of a particular face in memory, ii) knowing why the face is familiar and iii) recalling the person's name. This model does not apply to unfamiliar faces, and thus the context in which a person was encountered, and for how long, define the strength of the memory. The greater the number of encounters with someone, the more semantic information is available about them (Bruce, 1982), and even when the name of the person is not recalled, other available codes make the face familiar enough to be recognised, as can sometimes be the case with celebrities (Klatzky & Forrest, 1984). The unfamiliar face recognition's vulnerability to errors can also be understood by recognition relying on pictorial codes (i.e., properties of the encounter such as lighting at the time), rather than structural codes (i.e., information about aspects of facial identity; Bruce, 1982). For example, if identical images are seen of a person between a study (encoding) and test (retrieval) phase, recognition may be based on the same angle and expression of the face, or clothing, rather than genuine properties of the face related to identity, such as configural information, which are independent of these factors (Bruce et al., 1982; Bruce & Young, 1986).

1.3.2 The view of the face at encoding

A witness of a crime is exposed to a criminal's face for a different length of time but, in the vast majority of cases, not long enough for the face to become familiar. One of the tasks of a composite operator, or artist, is to find out the rough length of time and the view that the witness had of the face and proceed accordingly. In most cases, the composite will be of a frontal view, particularly when a computerised system is used—although composite systems, for example PRO-fit can construct the face from a three-quarter view (Ness et al., 2015). Other systems that can generate a 3D view are 3D face models (e.g., Blanz, et al 2006; Huang et al., 2003) and the researcher is aware of one pilot study involving using a 3D programme for composite construction (Johnson, 2010). The sketching method is more flexible at this task, as any view could be created (assuming that the forensic artist has the ability to do this). Proficiently this would seem to be helpful especially if a profile was the sole view the witness has had, or a sketch from that view may complement the frontal image. If an unfamiliar face has been encoded in various views, this usually aids recognition (e.g., Chen & Liu, 2009), and a three-quarter view has been found to be most recognisable (Baddeley & Woodhead, 1983; Bruce, Valentine and Baddeley, 1987; Bruce & Young, 1998; Krouse, 1981; O'Toole et al., 1998), particularly for unfamiliar faces. This is because such a viewpoint offers most information about a face (Hancock, 2012). The advantage of a three-quarter view requires that the face had been encoded at this view, and thus encoding a front view does not lead to a recognition advantage at three-quarter view.

Kelly Lawson, who works as a sketch artist in Georgia, USA, reports that she draws a sketch composite based on the best view of the witness's mental image of the perpetrator, which is often a three-quarter view (K. Lawson, personal communication, April 6, 2021). She also mentions that it can also be a face viewed from a slightly upturned or downturned view for example. While no empirical evidence exists for how effective such a sketch orientation would be, Lawson claims that many of her composites are successful in leading to an identity of the perpetrator. It is noteworthy that Lawson's craftsmanship is high quality (see figure 1.8), and sketches contain a lot of shading, making them highly realistic looking. This level of skill would appear to be rare (wording changed as suggested) and not all sketch artists would be confident enough in their skills to draw any view of the face. In the experiments carried out for this thesis, a frontal view of the face is drawn due to the stimuli being mostly front facing images (apart from Experiment 1, which used video clips); however. Sketching is capable of this possibility due to its flexible nature and has an advantage over computerised systems in this respect, provided that the artist is sufficiently skilled and confident. For the principles discussed above, and to mimic a real-life situation, it is important that the target faces used are also unfamiliar to participants who construct composites, as otherwise prior familiarity is likely to inflate accuracy in these cases (Frowd et al., 2011), adding noise to one or more conditions in an experiment.

Figure 1.8 Sketch composite by Kelly Lawson. Permission to use the image gained from Lawson (2021).

If an unfamiliar face closely resembles another person's face, this can make identification more difficult, and potentially lead to misidentification errors (Kemp et al., 1997; Young et al., 1985b). For example, when participants in Henderson et al. (2001) were presented with an array of eight similar looking males including another target used in a different array (the correct target was not present in this array), they wrongly picked a non-target 31 % of the time. Participants were shown a good quality image, which matched the wrong target's viewpoint and expression, and the images in the array were CCTV still images captured using broadcast quality video. The same difficulty is observed in real life. Viewing CCTV (closed-circuit television/security video) images of faces further demonstrates the difference between familiar and unfamiliar face recognition: familiar faces are recognised easily even from very poor-quality images, while recognition of unfamiliar faces poses problems even when image quality is very good (e.g., Bruce et al., 1999; Burton et al., 1999; Jenkins et al., 2011). Even if people have an ID card to compare the person to, they are still prone to either reject a correct identification or accept a wrong one (Kemp et al., 1997).

Resemblance to someone else can also be beneficial in some cases. If the person happens to resemble a celebrity for example, the semantic processes may strengthen the memory for that face (e.g., Hole & Bourne, 2010). During composite construction, the interviewer might ask a witness whether the person resembled anyone familiar to them, and this can potentially act as a cue to memory. In contrast, identification of a composite can be made only by people who are familiar with the identity. The main difference between recognition of an actual photograph of a person and a facial composite is that a facial composite will always contain error (Frowd et al., 2007), being based on eyewitnesses' memory. However, this need not matter to a great extent since constructed facial features, especially if they are distinctive (e.g., Light, Kayra-Stuart, & Hollander, 1979), may lead to someone in the public suggesting a suspect to the police as one or more features may have triggered recognition (e.g., Gibson, 2010; Taylor, 2001; also see Frowd et al., 2007).

1.3.3 Internal and external facial features

Areas of the face are also recognised differently depending on whether a face is familiar or unfamiliar to an observer. As a face becomes increasingly familiar, the expressive internal features become more important (Ellis & Shepherd, 1992), and hence the face tends to be recognised more by internal facial features (eyes, brows, nose, mouth), whereas for unfamiliar face recognition the external features play a more important role (e.g., Ellis, Shepherd, & Davies, 1979; Young et al., 1985). Indeed, we tend to find it difficult to recognise someone who has only been seen once if their hair is concealed or changed (Ellis et al., 1979), and this dominance of external features in unfamiliar face recognition is found in tasks that involve direct matching of unfamiliar faces—that is, without any memory load (Bruce et al., 1999). The progression from unfamiliar to familiar is not well defined but even a short but repeated exposure of a face may allow familiarity to be reached to some extent. Clutterbuck and Johnston (2005) found an advantage for the internal features (eyes, nose, mouth region), although weak, after only ten two-second exposures to a previously unseen face.

O'Donnell and Bruce (2001) investigated participants' ability to detect configural and local feature changes to internal and external features of unfamiliar faces, and faces to which participants had been newly familiarised. For this latter group, participants were familiarised with faces by viewing a 20 second video clip of a target's face rotating from profile to frontal view and looking up and down and smiling to the camera. Their task was to simultaneously learn the names until they remembered all the face-name pairs. One of the two faces in the image pair used in the test phase was manipulated: eyes and chin had their spacing (configural information) changed, whereas mouth and hair had this feature swapped with one from another face. There were six conditions in which the faces were shown: changed eyes, mouth, hair, chin, different person, and same person. In both experiments, hair change was most easily detected in untrained (unfamiliar) faces. In trained (familiarised) faces, detection of changes for eyes was enhanced, while sensitivity to hair changes were maintained. No enhancement was found on the other internal features. These results highlight the importance of hair for both types of face recognition, and eye region for familiar faces. How we process faces is of particular interest to understand how facial composites are constructed. The next section discusses the vast area of relevant research, holistic and featural face processing.

1.3.4 Holistic vs featural processing

Substantial research indicates that adult observers recognise upright faces in a holistic manner (e.g., Tanaka & Farah, 1993; Tanaka & Sengco, 1997), through attending to spatial relationships between features, rather than focusing on specific features themselves (Searcy & Bartlett, 1996; Tanaka & Sengco, 1997). It is easier to recognise any alteration in a single feature when the whole face is visible compared to when only a single feature is visible (Davies & Christie, 1982; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). And alteration of the spatial location of the eyes not only impairs the eyes but also other features that have their spatial location unaltered (Tanaka & Sengco, 1997). Famous faces have also been found to be more difficult to recognise when the top and bottom halves of faces come from a different person due to the new configural arrangement of the faces, in comparison to when the halves are presented separately (Young et al., 1987). Inversion of the faces makes it easier to recognise the halves from the whole face.

The whole-face process is usually disrupted when faces are seen upside-down (Yin, 1969), while other objects do not tend to suffer from a similar inversion effect (e.g., Bruyer, 2011; Bruyer & Crispeels, 1992; Yin, 1969), even for an object (e.g., a watch) that has a clearly defined internal and external feature structure (Meinhardt-Injac, 2013). Inversion leads to facial features being processed in a piecemeal manner, as is the usual case for other non-face objects (e.g., Young et al., 1987), which is not optimal for face recognition (e.g., Tanaka & Sengco, 1997). Researchers argue that configural processing is used to a greater extent when encoding faces with which we are familiar, while featural processing occurs to a greater extent when unfamiliar (e.g., Rakover, 2002; Rhodes, Hayward, & Winkler, 2005); however, see Hayward et al. (2005), for the own-race faces demonstrating both configural and featural processing advantages over other-race faces. Since faces are familiar stimuli in general, we are likely to rely more on configural processing to encode them (Farah, Tanaka, & Drain, 1995; Wilford & Wells, 2010), although a more inclusive review of the appropriate research indicates that both configural and piecemeal mechanisms are likely to be involved (see Bruyer, 2011). It has also been suggested that holistic processing is faster because facial features are processed in parallel rather than one by one, and is more reliable because individual facial features (e.g., skin shade, mouth shape) are subject to change, while the configural arrangement is relatively more stable (Eysenck & Keane, 2015).

Configural processing seems to be related to level of expertise in recognising a given type of object. For example, people with considerable exposure to non-human faces such as a specific dog breed (Diamond & Carey, 1986) can gain a similar level of expertise in recognising individual dogs as our ability in recognising human faces (Gauthier et al., 2014). Since most of us are exposed to human faces constantly in our daily lives on a general level and not to dogs of a specific breed, there is a tendency to call most people experts at face recognition (e.g., Carey, 1992; Carey & Diamond, 1977; McKone et al., 2007; Tanaka & Gauthier, 1997). However, this so-called expertise is likely to be more on a general level of face processing, linked to the level of familiarity of the face since humans are not very good at recognising unfamiliar faces, as mentioned earlier (e.g., see Young & Burton, 2018). Our memory does not contain a model for an unfamiliar face like it does of familiar faces (Hay & Young, 1982).

The inversion effect is not evident in the same way in young children than in adults (Carey & Diamond, 1994), or in faces from a less familiar racial group compared to faces from the same racial group (Rhodes et al., 1989; but see Valentine, 1991). This is in comparison to the robust finding that same race faces are recognised better than other race faces in a normal upright orientation (Valentine, 1991); and that young children have yet to reach this more adult level of performance (e.g., Chance, Turner, & Goldstein, 1982; Goldstein & Chance, 1980), indicating that their face recognition is at a more basic level rather than having proceeded to a more specific subordinate level, one that defines a person's identity to a greater extent (e.g., Tanaka, 2001). Goldstein and Chance (1982) suggest that children become more efficient at using a face schema with age, making them better at recognising faces in general. Valentine (1991) suggests that, as a side effect, their face schema also becomes more rigid, in turn making them worse at recognising inverted and other race faces. It has been suggested that this is because young children do not have as much experience as adults in seeing upright faces, and therefore lack expertise (Tanaka, 2001). Configural processing of a face has been found to develop more slowly than featural processing, and children were almost as good as adults at featural face recognition, while significantly worse at configural processing (Mondloch et al., 2002).

More evidence for holistic processing of faces can be seen in how caricatures affect recognition. Caricatures, which exaggerate facial features and proportional information from an average reference face, are often recognised better than the real photo of the face, which suggests that the encoding of faces occurs by such deviations (e.g., Rhodes et al., 1987). The impact of dynamic proportional relationships is demonstrated by Webster et al. (2004), who found that by concentrating on a face whose features deviate from average and then looking at the average face leads to briefly distorting the average face, a finding (of adaptation) that supports the above theory. In the current context, a dynamic caricaturing effect, in which a composite is seen moving from negative and positive caricature states, has been found to improve recognition of composites created from different systems (e.g., E-FIT, PRO-fit, EvoFIT and sketch). This sequence is believed to be effective by reducing error in composites during the negative caricature states, while enhancing distinctive information in the positive caricature states (Frowd et al., 2007, 2012).

It has been debated how constructing a facial composite affects later recognition or identification of the target. Focusing on individual facial features, for example, by constructing a composite with feature systems such as IdentiKit has been shown to improve face recognition (Mauldin & Laughery, 1981; Meissner & Brigham, 2001), but may also interfere with memory (e.g., Comish, 1987; Wells, Charman, & Olson, 2005). This may allow different, less deep (less holistic) processing for the memory for the target, and yet it may improve memory for facial features. There has been much research in the area, some of which has investigated the impact of face construction on identification (for reviews, see Sporer et al., 2020). Some research has generally indicated positive benefit (e.g., Sporer et al., 2020; meta-analysis by Meissner & Brigham, 2001) and occasionally negative (e.g., Wells et al., 2005). In Davis et al. (2014), for example, construction of an EFIT-V (holistic) composite and an E-FIT (featural system) composite by participants enabled them to more accurately identify a face in a line-up (cf. no construction of a composite), and for EFIT-V even with a more realistic 72-hour delay from target encoding to line-up viewing (although the composite was constructed following a very short delay after target encoding). In a second experiment, participants constructed either one or three EFIT-V composites. In both experiments, participants were one-and-a-half times more likely to make a correct target identification from a video line-up than those who did not construct a composite at all. This research (along with others, Meissner & Brigham, 2001) indicates that face construction in general is beneficial for retrieval of facial memory.

Achieving accurate facial configuration may be challenging to achieve in a sketch composite and is a reason why it is important to investigate different methods likely to achieve a good match for this aspect of the face. Such an objective could be achieved, for example, by using reference materials after free recall, or by creating a sketch based on recall and use of probing questions to elicit more information. Since constructing facial composites is a fundamentally different procedure depending on the system used, how these methods compare to each other is discussed next.

1.4 Development of the face construction procedure

1.*4.1 From face recognition studies to investigating recall*

While research has been focused on recognition studies, face recall has not been similarly researched due to the lack of an appropriate tool being available for quantitative research until the mechanical composite systems were developed (Ellis, Shepherd, & Davies, 1975). Ellis et al. (1975) appears to be the first laboratory study investigating face recall through a construction of a facial composite with Photofit. They used two Caucasian male faces as stimuli. The participants (N = 34) constructed both faces using Photo-fit (the first face from memory after it had been shown to them for 10 seconds and the second face while the target was present). The composites were evaluated by 12 independent judges who used two methods:

1) absolute number of correct features selected, where each correctly selected facial feature was scored 1 point, making 5 the best score for the five features (forehead, eyes, nose, mouth and chin) used to construct the face on Photofit; and 2) goodness of likeness ratings on a 7-point scale: 1 (poor match) – 4 (moderate match) – 7 (perfect match). The participants compared each composite with their target Photofit face, and the order of the composites was randomised. The reliability of ratings was evaluated to check if the judges' ratings varied, and the judges were found to be in close agreement with each other. There was no significant difference in ratings between the two target faces; however, the ratings are much higher when the composite was constructed while the target was present as opposed to from memory. Thus, they confirmed that a significant difference was found when data of the two target faces were combined, although even from view the composites were far from perfect likenesses. The possibility of constructing an exact likeness was present, since Photofit itself was used to construct the two Caucasian male target faces used in the experiment. The results were likely affected by the fact that the constructors were members of the public who had received no training prior to the study and the process was only explained before the first construction.

In Experiment 2 by Ellis et al. (1975), six faces out of 36 black and white images of Caucasian males were randomly selected to be constructed with Photofit by the six previous participants who had scored highest in the memory condition in Experiment 1, and the six participants who had scored the lowest. Each participant constructed three of the faces in the same way as before from memory. The faces were sorted into six sets of six so that each set contained three Photofits from the higher scorers and three of the lower scorers. A composite of each of the six original faces was contained in each set of slides. Seventy-two new participants were put into six groups of 12. One slide was shown to each group in a random order. They were told they would see one Photofit and 36 other faces in a 6 x 6 matrix on another screen and instructed to choose which one looked most like the Photofit. They were also asked to pick second and third choices. 1 in 8 (12.5 %) of the 72 judges chose the correct face, which is obviously quite low but above chance (cf. 1 in 36). Incorrect answers were given by 1 in 4 (again, above chance level: 1 in 12). Thirty five of the 54 correct answers were Photofits of the better scoring participants, and thus significantly more Photofits of the higher scorers were matched to the correct face than the Photofits by the lower scorers. These two experiments indicated that there were individual differences between people who constructed composites, and participants who produced better Photofits in Experiment 1 also constructed better composites in Experiment 2. It was also concluded that Photofit was limited in its technique and would benefit from more features being available. It is likely that some people are intuitively better at composite construction and interacting with witnesses, but experience is also likely to play a part. This study forms a basis for using likeness ratings in composite studies.

From the 1970s, facial composite research has been increasing, focusing mainly on mechanical systems. As with Ellis (1975), these studies tended to use likeness ratings and/or sorting tasks as a means of evaluating the effectiveness of composites (e.g., Ellis et al., 1978). The results indicate that the mechanical systems such as Identi-Kit and Photofit were not performing well for producing composites that closely resemble the intended subject. Such systems have been found to have technical shortcomings. Specifically, it has been found that there were not enough facial features from which to choose (e.g., Laughery & Fowler 1980), features were restricted to Caucasian faces (Kitson, Darnbrough, & and Shields, 1978), and there was a “clunky” composite editing method involving a pen and transparency film (Davies, Milne, & Shepherd, 1983). These systems have also been found to be less effective than verbal descriptions on their own (Christie & Ellis, 1981), which indicates that recall itself may not be the problem, rather the visual interpretation was being produced by an inadequate tool. Yet, in the absence of a better system, police around the world continued to use these systems, as sometimes they would prove useful in an investigation (Kitson, Darnbrough, & Shields, 1978). In Britain, for example, Photofit was still used widely in the 1990s before being replaced by computerised feature systems such as CD-FIT and E-FIT (Gibling & Bennett, 1994). More fundamentally, the mechanical systems do not tend to match with how faces are processed in a more holistic manner (e.g., Tanaka & Farah, 1993; Tanaka & Sengco, 1997), and this, along with the limited technique, render these systems ineffective.

Sketch artists were employed in Britain and North America (US and Canada) along with operators for mechanical systems between 1960s and 1980s (Davies, 1986; Jackson, 1967). However, artists were not an obvious choice for creating composites in the early research projects, and Ellis et al. (1975) make no suggestion for including them in future research. The fact that, in Britain, they worked indirectly with witnesses, would not have harnessed the benefits of sketching and sketches were thus perhaps not considered as reliable as mechanical systems; and as such, this is likely to have led to some unfavourable opinions being formed about sketch composites. In the US, being a police artist seems to have been more of an established, albeit niche, career; sketch artists were even included in laboratory research by Laughery et.al. (1977), as described. Since relatively few people have the skills to interpret a witness’s descriptions and formulate it into a realistic looking sketch, the use of artists appears to always have been in the margins of forensic investigation.

Sketch composites have been claimed to rarely match accurately with actual photographs of the suspects, especially when automated matching of sketches to a database of mugshot is used (e.g., Klare, 2011), although improved automated matching has found more favourable results (Kokila, Sannidhan, & Abhir, 2017). These studies focussed on the measurements of the face though, and a facial composite can be successful even if some features or even proportional information contain inaccuracies. Kokila et al. (2017) do take this issue into consideration to some extent by allowing more error in the distances between facial features. An image that has captured the essence of the offender or includes a distinctive feature that is emphasised, can be enough to trigger familiar face recognition in someone who knows the subject (e.g., Gibson, 2010; Taylor, 2001). This is evident in studies using caricatures, which have been thought to act as ‘super-portraits’ (Hole & Bourne, 2010). Positive caricatures (distinct features exaggerated) have been found to be recognised better than the original faces from which they are derived, while anti-caricatures (distinct features altered to be more average) have had the opposite results (Benson and Perrett, 1994; Rhodes et al. 1987). For photographic images this effect has much smaller effects on the speed or accuracy of recognition as opposed to drawings, and these caricatures are also rated as poorer likenesses of the faces concerned than their undistorted versions (Benson and Perrett, 1991; Kaufman and Schweinberger, 2008). This is possibly because line drawings have less texture information than photographs, giving more scope for caricaturing to produce an effect with drawings (Hole and Bourne, 2010), as the focus is on shapes of features represented in a more simplistic way. This indicates that measuring sketch composites based on the absolute accuracy of the proportions of the facial features is not a sufficient way to measure their utility in leading to someone recognising a composite face as someone they know or have seen.

Therefore, it is also important that the police officers/staff and the general public understand that composites are not exact likenesses but are on a “sliding scale” of likeness and are likely to contain inaccuracies due to being depicted from memory. Even if a witness had a good memory and a clear mental image of the offender, verbalising detailed information for another person to interpret as a visual representation of their memory is still a challenging task and can contribute to inaccuracies in the resulting composite. It is obviously possible to achieve a good likeness to the suspect with this technique; there are many examples from real cases where a sketch composite has been effective in generating a lead to the suspect’s identity (see Boylan, 2000; Gibson, 2010 and Taylor, 2001 for examples). A question remains, though, how helpful sketches are overall in criminal investigations. Many sketch artists report anecdotally that it can be difficult to keep track of the ensuing identification of composites, since some investigations may take years to complete and the officers do not always update the sketch artist of the outcome, even if the sketch did help in the identification of the offender (Enslow, Falsetti, Lawson, Zamora, personal communication, 2011-2022).

1.4.2 Sketch composites in laboratory studies

The first study to include sketch composites in research appears to be by Laughery, Duval and Fowler (1977). This study compared sketch to Identi-Kit composites. The experiments employing sketch artists were part of a Mug File Project in which facial images produced in their experiments provided a database for pattern-recognition algorithms. The main aim of their project was to develop an interactive computer system for criminal identification, but the experiments provided valuable information on mechanical and sketch composites. The authors went to considerable lengths in designing the methodology of the experiments and describing the sketching process, the latter of which will be considered in more detail later in the section below (focusing on the sketching technique). The sketch artists employed in the studies were art graduates with a good level of experience in portraiture, had practised drawing faces from descriptions extensively before the experiments, and two of the artists had produced sketches with witnesses for the Houston University Security Office and one for the Houston Police. It was stated that in a forensic setting, a standard identification procedure is to show the witness a large set of mug file photographs to search for suspects. The mug file is first narrowed down based on a witness's verbal recall before she or he goes through the images and determines whether the suspect can be identified among the mugshots. The objective of this project was to emphasise the witness's recall by constructing a facial composite and for the computer to locate look-alikes from mugshot files. Three experiments were completed on different target populations, white males, black males and white females. Experiment 1, using white male targets, employed three artists and three Identi-Kit technicians, and is in fact the same experiment described in Laughery and Fowler (1980). The two other experiments presented, which involved black males and white females, were essentially the same as Experiment 1, although less data were collected. Two artists and two Identi-Kit technicians were employed, with 20 targets and 40 witnesses in each study generating one of each composite. The experiments were trying to emulate a real-life situation as the participants, who were students from the University of Houston or members of the larger community in Houston, met a target in pairs, had a conversation with him or her and were then asked to construct a composite either with a sketch artist or an Identi-Kit operator. The target was a different face for each pair of mock witnesses, an unknown identity to them and to the composite constructors. The artists and Identi-Kit operators also constructed the targets from view after the memory condition. The first 'goodness of fit' measure of the composites was a likeness rating task carried out by an independent set of participants.

Sketches were rated as better likenesses to the target overall than Identi-Kit composites. The benefit of having the target in view (higher likeness) while constructing a composite was also only evident in sketches. These results indicate that Identi-Kit had limitations in its technique while sketching is a more flexible technique, and the sketch artist was better able than an Identi-Kit operator to apply alterations to the face as required by the witness. A second dependent measure used to assess the goodness-of-fit of the composite images was based on ten physical measures of the images and faces. A computer algorithm, developed as part of this study, was then used for selecting look-alikes from a mug file. Sketches were ranked better than Identi-Kit composites, in both conditions, which is consistent with the outcome from likeness ratings. However, there was no difference between in-view and memory for this DV with sketches, unlike in the rating task.

The superiority of sketches is likely to indicate that the proportional information can be more easily depicted using this method, as there are no limitations where to position facial features (cf. Identi-Kit). Taylor (2001) emphasises the importance of proportional accuracy in achieving a likeness in sketch composites. Laughery et al. (1977) found that witnesses tend to move between facial features to a greater extent when a sketch is being created compared to working on an Identi-Kit composite. They suggest that it may be that sketching results in better relationships (e.g., distances) between features than when the focus is on one feature at a time. This seems to indicate that sketching is midway between a feature system and a holistic system but may change depending on whether reference materials (pictures of facial features) are used in the process. Reference images that show one feature at a time by having a shape blocking the rest of the features, such as in some facial feature catalogues that artists use (e.g., Steinberg, 2006), intend for a pictorial reference to jog the witness's recognition memory without causing too much interference to the mental image by seeing whole faces. While many sketch artists advocate the use of these type of references (e.g., Mahoney, 2010; Steinberg, 2006), others prefer to use whole faces and do not think they cause much interference to facial memory from their experience (e.g., K. Lawson, personal communication, April 6, 2021; Mancusi, 2010; Taylor, 2001). An argument for the latter technique is that this way of processing faces is more holistic than seeing individual features, which has been found to lead to better recognition of faces (e.g., Tanaka & Farah, 1993; Tanaka & Sengco, 1997).

Laughery and Smith (1978) also compared sketch against the earlier version of Identi-kit that contained line drawings rather than using the updated version containing photographic images. According to Shepherd and Ellis (1996), no formal evaluation of this latter version of Identi-Kit exists. In Laughery and Smith (1978), it was found that sketches that were rated high in similarity to the target photograph were recognised at accuracy rate of 71% from mug shot files, while for Identi-Kit, the recognition rate was 51%. Previous research indicates that line drawings do not tend to be effective for recognising well-known people (e.g., Bruce et al., 1992; Davies et al., 1978; Rhodes et al., 1987), even when there is an accurate representation of the face, and fine details such as facial lines and wrinkles are included in the face (Davies et al., 1978). Davies et al. (1978) and Bruce et al. (1992) suggest that the problem might be a lack of three-dimensional shape represented in the image that gives more structural information to faces, aiding recognition. This deficiency could contribute to a lack of resemblance caused by a difficulty to create distinctive faces, which has been found to be difficult to represent compared to less salient features (Green & Geiselman, 1989). It is established that facial distinctiveness is an important factor for face recognition (e.g., Shapiro & Penrod, 1986), as well as for accurate naming of facial composites (Frowd, et al., 2005).

Identikit has since been revised, and a version now exists called Identikit 2000 that is a computerised sketch-like feature system. It is used widely in the US, and evaluated against PRO-fit by Frowd, McQuiston-Surrett, Anandaciva et al. (2007). The correct naming results for both systems were very low, and a novel “cued naming” task was relied upon as another method of assessment. Using this procedure, participants first attempted to name composites, then named target photographs, and then attempted to name the composites for a second time - called cued naming as the composites were “cued” by having seen the target identities. A supplementary task, sorting, was also included. This involved participant matching the composites to their target photos. Both evaluation tasks revealed that these two systems performed equivalently.

1.5 Adding flexibility to mechanical systems by artistic skills

1.5.1 Photofit-sketch hybrid method

As mentioned above, in the UK, the usual way of producing sketch composites was for an artist to work indirectly, that is from witness descriptions provided to them along with an existing Photofit. A British sketch artist Melissa Little (M. Little, personal communication, July 16, 2021) has suggested that artists were sometimes employed to make alterations to an initial Photofit composite if no appropriate feature(s) was found in the Photofit kit, or if the composite needed further alteration. She believes the indirect method of working with a witness risked misunderstanding occurring more often than if direct interaction with a witness had taken place. No evaluation of this hybrid Photofit-artist technique existed, and so Davies (1986) set out to compare it to the conventional Photofit process which was being produced with witnesses in a study by Christie et al. (1981). Davies (1986) employed a professional sketch artist who had sketched composites from descriptions on several occasions. Also, a line transcription of a Photofit was included as a third condition to test a hypothesis that a more schematic, less realistic drawing would lead to higher identification ratings, especially when likenesses were poor. The Photofits had been produced after a one-minute exposure to a still image of the target face. Each of these 36 composites were evaluated by another set of 10 judges who were asked to find the target from an array of 24 photographs of young men based on the composite. As with Laughery and Smith (1978), the most accurately identified (“good”) and least accurately (“poor”) identified composites of each of the 6 targets were selected as stimuli for this study. The ID rates for the good composites were 55% and poor ones were only 3%. Three sets of 12 faces were used as comparison stimuli. The first set, photofits, were simply monochrome prints of the 12 original photofits. The second set, artists’ impressions, were sketched by using the same photofit composites together with their associated verbal descriptions following normal police procedure. The 12 three-quarter view pencil sketches were then photographed and reproduced as monochrome prints. The third set, line transcriptions, were produced by taking each Photofit picture and projecting them on to a flat surface. The outline of major facial features was then drawn, using approximately the same level of detail as that of the original Identikit. Each of the 12 stimuli were then divided into two subsets representing the six good and six poor likenesses of the targets.

A new set of judges, students and the members of the public, evaluated the composites. Each judge was given a subset of six faces and a loose-leaf booklet containing 24 coloured mugshots, six of which were the targets and the remaining 18 were foils. Participants were instructed to search through the booklet for each given face. A total of 102 participants carried out the task, 17 being allocated to each of the six separate subsets of stimuli. The results revealed that, for the ‘good’ composites, the original photofits were significantly better recognised than either the line drawings or the artist’s impressions. For the poor composites, no significant difference was found. However, this might have been due to floor-level recognition. The incorrect choices were significantly lower for original Photofits than the other two groups for good composites, but, for the poor composites, there were no reliable differences.

These results mirror the accurate recognition rates. Applying the alterations from written descriptions seems a rather subjective approach when the witness is not reviewing changes, changes that might not have been interpreted correctly by the artist. The benefit of working directly with witnesses is that even small changes that are made to the sketch can lead to the witness being more confident with the likeness, and these changes may then prompt further changes, improving the likeness further. In Davies (1986), this interaction with witnesses did not occur and changes were applied only once based on the verbal descriptions of the original composite constructors, and thus this study is not a true evaluation of the sketch artists' technique used today.

Sketch composites in earlier laboratory studies have employed art graduates (e.g., Laughery & Fowler, 1980), and they have mostly been constructed immediately after target encoding (e.g., Laughery et al., 1977), or a sketch artist have been compared to art graduates drawing a face from their own memory (Davies & Little, 1990). All these scenarios lack realism with respect to real crime situations, and little empirical evidence exists on the utility of sketch composites (Davies & Little, 1990; Davies & Valentine, 2006), highlighting the need for a more robust methodology to understand the effectiveness of artist's sketches used in modern criminal investigations.

1.5.2 Editing options for mechanical composites

Due to their poor performance in representing the target face (see Shepherd & Ellis, 1996, for an excellent review of these systems), it is clear that mechanical systems do not match the more natural, configural way of processing faces, in part because the focus of attention is on isolated facial features (e.g., Christie and Ellis, 1981; Davies and Christie, 1982). In an attempt to add flexibility to face construction, editing the initial composite was made available using another transparent sheet placed above the assembled features. The effectiveness of applying alterations to initial composites was measured on Photofit (Davies, Milne, & Shepherd, 1983). When Photofits were constructed from memory (immediately after a one-minute encoding of the target face) by an expert operator and a novice, added artwork did not improve the effectiveness of the composites, and it was suggested that the superiority of the expert's performance was achieved in the initial stage of composite construction. However, Photofit composites constructed by the experienced operator were evaluated as better likenesses than the novice's composites overall, which is in contrast to the earlier studies on these systems and indicates that operator experience plays a part in the quality of composites, and even a system with technical limitations is likely to be more flexible in the hands of an expert. The option to make alterations to the composite by drawing on a transparent film would have been expected to make the use of Photofit more flexible but, since pen would have been applied to a photographic image, this might not have been an efficient way of editing the composite.

In contrast, Gibling and Bennett (1994) found that when Photofits were constructed while the target was in view, enhanced Photofits were of superior quality and better likeness than unenhanced ones and led to significantly increased identification accuracy. They employed student Photofit operators undergoing training at the time and each constructed one Photofit composite. Three methods of artistic enhancement were introduced. First, two acetate sheets could be used to first blank out the unwanted areas of the initial Photofit image and on the other sheet alterations could be made. Neck, shoulder and clothing could also be added in this way. Second, if more enhancements were required, the Photofit could be photocopied, and that representation could be worked on further. And finally, the correct Photofit features could be traced, and this could be used to produce a sketched impression of the face. They compared the original Photofits and Photofits enhanced by using pencils and pastels on a matt-lustre acetate placed over the original composite. Six photographic line-ups were constructed for each target face and police officers attempted to identify the target faces from the composite. The enhanced Photofits were found to be significantly better identified than the original ones, indicating that artistic enhancement does make a positive difference to the likeness of the composite, at least in non-forensic conditions where the target face is visible. In Davies et al. (1983), the expert spent more time constructing the composites with witnesses than novices when they worked alone instead of in pairs. This indicates that the initial composite was constructed more carefully through having more experience and perhaps since the Photofit operators in Gibling and Bennett (1994) were still in training, they did not spend quite as much time on this stage and thus additional artwork was more effective. The lack of evidence of any image editing benefitting the composite quality, when the composite was constructed from memory, suggests that the problem is mainly in the limiting process of constructing the initial composite face by the restricted number of facial features that could be selected. After all the features have been selected, the process may interfere with the constructor's mental image of the target, and thus it might be too late for the artistic editing of the image to benefit the outcome of the composite.

1.5.3 From mechanical systems to computerised systems

The early mechanical systems would be considered rather old fashioned and clunky compared to the computerised systems that started to appear in the 1990s. It has perhaps been assumed that a larger bank of facial features to choose from and more sophisticated ways of editing facial features and their proportions, firstly within the computerised system itself and secondly by transferring the initial composite to an image editing programme, would improve composite likenesses. Although there is some indication of more flexibility in the computerised system, leading to better composite likeness when the target is in view while a composite is being constructed (e.g., Cutler et al., 1988), the desired result has largely not been found with the computerised feature system (e.g., Koehn & Fisher, 1997). A study by Davies, Willik and Morrison (2000) was the first to compare the computerised E-FIT and the mechanical Photofit. The participants were exposed to the target image (one of four university members of staff) for 1 minute and the composites were then constructed. The composites were evaluated by a naming and a sorting task by further participants. The correct naming of the composites was low overall (17%), and no significant differences were found in either evaluation task. Only E-FIT performed significantly better when composites were constructed with the target in view, indicating that a computerised system has greater ability compared to mechanical systems; however, the computerised system still performed poorly, to the same level of naming as the previous mechanical system for the intended use, when a face was constructed from memory. Frowd et al. (2000; 2005) achieved similar level of correct naming with E-FIT, and Bruce et al. (2002) with PRO-fit.

1.6 More attention to the interview

The importance of the initial interview and how detailed information is drawn from the witness, is emphasised with feature composite systems including sketching, especially if witnesses find recalling the face particularly difficult. Research in the early composite systems tended to focus mainly on the technical aspects, and not the interview. These research projects were conducted before the ground-breaking cognitive interview (Fisher & Geiselman, 1992), and so may lack the benefit afforded by the aspects of this interviewing technique. However, guided memory, a technique similar to mental reinstatement of context (a mnemonic of the Cognitive Interview) prompts a witness to recall the circumstances and the environment of the crime, was found to lead to more accurate information when Photofit composites were constructed (Davies & Milne, 1985).

Davies and Milne (1985) appear to be the first to investigate the importance of context in composite construction using a more realistic, incident methodology. Participants observed one of four targets walk into a room searching for a calculator for approximately 1 minute 15 seconds, and after a one-week delay returned to construct a composite using Photofit. The one-week interval was based on the realistic scenario (at the time) where most composites are made within a week of an incident (Darnbrough, 1977). Participants constructed a composite either in the same room as where the target had been seen or in a different room and were given either a guided memory procedure modelled on that of Malpass and Devine (1981), or without this guidance (spontaneous recall). Participants in the guided memory condition were asked to think back to the event, the environment and their mood and emotions at the time. The focus of the interview then switched to the appearance of the target, firstly the general appearance and then to construct the face.

The resulting Photofit composites were evaluated by another group of participants who attempted to match them to photographs of the four targets. It was found that the composites created in the guided memory condition were significantly more recognisable (i.e., were matched more accurately) compared to those from the spontaneous recall condition, while a less effective result (although still significant) was observed for face construction in the same room as encoding. It was suggested that there was less of an effect of the environment in the guided memory condition, but the interaction of the factors was not significant. The composite evaluators were supplied with the relevant verbal descriptions of the target as well as the composites, which could have supported the recognition of the composites in some cases; however, there was no marked difference in the length of the descriptions across conditions, and thus it was concluded that results were mostly due to the effect of the composite. These findings support the value of the mental reinstatement of context mnemonic, which can be even as effective as being in the same physical environment at encoding and retrieval stages (Smith, 1979). Given early evidence that aspects of the interview can improve composites constructed with mechanical systems, more focus on enhancing recall in facial composite construction appears to be a sensible step in improving the effectiveness composite systems. For this reason, the next section discusses memory and the Cognitive Interview in more detail.

1.7 Retrieving witness memory

1.7.1 Storing memories

A witness's account of an event of crime requires detailed retrieval of an episodic memory, which consists of an individual's personal experiences and involves them either as one of the acting members in the event or as an observer (Tulving, 1972; 1983). It is a process that enables the individual to 'travel back' mentally into his or her personal past (Tulving, 1998,2002). Only episodic memory supplies contextual information, rendering information specific rather than general (Surprenant & Neath, 2009). Episodic memory is thought to have unlimited capacity (see discussion in Capaldi & Neath, 1995) and that forgetting is due to interference (Surprenant & Neath, 2009). This system is thought to be synonymous with conscious awareness and it is evolutionarily and ontologically the highest form of memory (Sherry & Schacter, 1987; Tulving, 2002). Episodic memory should develop according to a different time scale than other memory systems. In addition, it can be destroyed without disturbing other more primitive memory systems. It is usually tested with recognition or recall of previously presented information (Surprenant & Neath, 2009).

Contrary to episodic memory, semantic memory consists of general knowledge about the world (Tulving, 1972; 1983); however, these two types of memories share more similarities than first thought (Tulving, 1983). Retention in both systems is thought to be automatic and the information is brought to awareness by a retrieval process (Tulving, 1983). When a composite is constructed, tapping into semantic memory may provide useful cues for recalling an offender's face, for example if the witness states that the face resembled another person she/he is familiar with, a celebrity perhaps. Both episodic and semantic memory require verbalising information to other people, while procedural memory consists of skills and procedures (for example driving a car) does not because this type of memory has been repeatedly practised (Tulving, 1983; Tulving & Craik, 2000), and is partly why constructing a facial composite presents additional challenges with interpretation of the mental image the witness holds.

The same regions of the brain are involved in the different memory systems. However, these systems interact with each other and share information (Tulving & Craik, 2000). There are areas in the brain that are specifically responsible for face perception. These include the fusiform gyrus (e.g., Damasio, Damasio & Van Hoesen, 1982; Kanwisher, McDermott & Chun, 1997), which was found to be activated in episodic memory both at encoding and recognition using pictures (Vaidya et al., 2002). Other areas include the inferior occipital gyrus or OFA “occipital face area’’ (e.g., Gauthier et al., 2000), and an area in the posterior part of the superior temporal sulcus (pSTS) (Puce et al., 1998). When a facial composite is constructed, the focus is naturally on the facial appearance of the offender. The face is, however, not separate from the rest of the event and thus, while we are interested primarily on depicting the face (and sometimes clothing and accessories too), any memories connected to the event, where the offender was seen, potentially enhance memory for the face.

1.7.2 Cognitive Interview

Recalling events is a cognitively challenging task for a witness. The strength of the memory depends on many factors: distance between the witness and the offender, lighting and weather conditions and angle of view being key elements for a memory for facial detail to have formed. These factors are referred to as estimator variables (Wells, 1978), factors which cannot be controlled by the justice system, or have already occurred (i.e., they are in the retrieval stage during the encoding phase). In addition to these external factors, there are numerous internal factors that have an impact on the accuracy of memory retrieval such as a witness's arousal level, attention to detail and the level of stress experienced. Fisher and Geiselman (1992) state that forgetting can occur even if the memory of the event has been stored. Retrieval of the episodic memories can be supported by an appropriate interviewing technique, one that aims to facilitate the witness's memory retrieval in the best possible way so that the elicited information is as elaborate and accurate as possible. Retrieval cues are used to aid the witness's memory, and the way that information has been encoded and stored determines which cues will be effective (Tulving, 1974). It is important to remember that the occurred event in memory is not an exact replica of what happened but rather how our minds have encoded it, since selective encoding omits some of the finer details (Quas, 2000). The role of the interview process is even more important to sketch composites than it is to computerised systems, as sketching relies more on face recall.

Given a lack of guidance for police officers to interview eyewitnesses effectively so that more accurate information would be elicited of the events, it was found that interviewer behaviour in fact hindered a witness's retrieval process (Dando & Milne, 2009). This led to the development of the Cognitive Interview by Geiselman and Fisher and colleagues *(*Geiselman et al.*,* 1984; Geiselman et al., 1985, 1986) in the early 1980s. Drawing from literature on cognitive psychology to identify optimal techniques for enhancing memory retrieval, the original Cognitive Interview (Geiselman et al., 1984) included four retrieval components for a witness: report everything; mentally reinstating the context; recall events in different temporal orders; and change of perspective. When a witness is recalling additional information, three interviewing techniques are critical for maintaining relatively high accuracy: (i) the witness is told not to guess or fabricate answers, (ii) use of open-ended questions and (iii) minimal use of leading questions. These effects were understood prior to the development of cognitive interview technique (Fisher et al., 2000). For example, the wording of questions has been found to influence what witnesses recall (e.g., Loftus & Palmer, 1974). Other effective interview methods include Conversation Management, the Memorandum of Good Practice, and the Stepwise method, all of which share the same core elements as the CI (Geiselman & Fisher, 2014). This thesis focuses on the cognitive interview due to its rigorous testing and it being the recommended interview technique when constructing facial composites (Association of Chief Police Officers, 2009; Richardson et al., 2010).

The effectiveness of the original Cognitive Interview was measured by Geiselman et al. (1985). Participants were shown a film and then interviewed using a standard interview (usual police interview technique, in which the police officer leads the interview in a series of questions and answers), a cognitive interview or a hypnosis interview. It was found that the CI outperformed the standard interview and was as good as hypnosis. While hypnosis appears to have generated mixed results, sometimes enhancing recall but sometimes not having any more impact than the control interview (see Orne et al., 1984; Reiser, 1989; Smith, 1983, for reviews), the cognitive interview seems to have consistently elicited more useful information than other investigative interviewing methods such as the standard police interview or a structured interview (Fisher et al., 2010). Several other laboratory studies by Geiselman and Fisher (Fisher, 1995) employed a similar design of college students witnessing a simulated event on a video and then being interviewed two days later. Across these experiments, the cognitive interview led to 25%-35% increase in more correct information compared to the standard police interview, and no increase in the proportion of incorrect statements (Fisher, 1995).

It was realised that more emphasis was required in the social interaction between the interviewer and the witness to facilitate recall, and this goal led to the development of the Enhanced Cognitive Interview (ECI) (Fisher et al., 1987), a protocol that considers the importance of firstly establishing rapport, to allow a witness to feel more comfortable in an interviewing situation. It is also emphasised that the interview process is explained clearly so that the witness knows what to expect and understands that the interview is a team effort rather than the interviewer bombarding the witness with many questions. The interviewer should also let the witness proceed at their own pace without being interrupted, particularly during free recall and context reinstatement, and let the witness review his or her recall, by for example repeating the witness's account, which can lead to the witness elaborating on it further. In addition to social dynamics, the interview is recommended to be conducted in an environment with minimal distraction. These changes to the original version of the Cognitive Interview allowed the interviews to become more relaxed for the witness since they address the interview situation more effectively (Dando & Milne, 2009).

Fisher et al. (1987) compared the enhanced cognitive interview with the original CI version and found that the ECI elicited 45% more correct information than the original version, with no significant decrease in accuracy. With the standard police interview, the benefit of the ECI doubled, even though the standard interviews were conducted by experienced police officers and the ECI by students who had received approximately 10 hours training in the technique. Also, the ECI has been proven to be more effective than other types of interviews, such as those used by social workers in child interviews (e.g., McCauley & Fisher, 1995). Köhnken et al. (1999) measured the cognitive interview's effects on correct and incorrect recall in a meta-analysis. This analysis included 42 studies, 55 individual comparisons and nearly 2500 interviewees. In 53 out of 55 experiments the cognitive interview outperformed the alternative interview in the amount and quality of information elicited. A large effect size was revealed for the increase of correctly recalled details with the cognitive interview compared to a control interview; however, also incorrect details were found to increase although with a considerably smaller effect size. Interestingly, the enhanced cognitive interview produced more errors than the original CI. Fisher (1995) points out that due to the multiple factors affecting eyewitness memory in a real-life scenario, which cannot be controlled similarly to laboratory conditions, we should not expect identical performance from these different situations. However, the experimental manipulations should in general replicate across situations (Fisher, 1995), making laboratory research valuable.

Laboratory studies have also highlighted the inaccuracies of eyewitness memory; however, field studies indicate that real victims and witnesses can be extremely accurate (Fisher et al., 1989; Yuille & Cutshall, 1986). Some evidence of this can also be seen in the aforementioned meta-analysis (Köhnken et al., 1999), as it was found that more correct details were elicited after the participant had encoded a staged event as opposed to video clips, and if they had actively participated in the event.

In Fisher et al. (1989), two groups of experienced detectives with equivalent interviewing skills from the Miami police recorded interviews with witnesses, with one group conducting the interview as per usual and one after receiving training on the cognitive interview technique. Training resulted in eliciting 48% more facts than prior to training and the trained group also elicited 63% more information compared to the untrained group. Clifford and George (1996) conducted a similar study with British Police investigators and found that their questioning style changed dramatically after having been trained in the cognitive interview. This included asking more open-ended questions, and fewer leading questions and provided more pauses, giving witnesses more time to answer. This change in interviewing style correlated with the amount of information elicited, which was higher for individuals after training and for trained investigators compared to untrained, similarly to Fisher et al. (1989). This advantage applied to a variety of types of information including details about the person witnessed. One downside of the field studies compared to laboratory settings is that the accuracy of the reported information cannot be measured; however, the laboratory studies support these findings.

The guidelines of the Cognitive Interview considering all the above research were set out by Fisher and Geiselman (1992), which is based on three psychological processes: social dynamics, memory and cognition, and communication. The witness can be a person who saw the suspect committing the crime or they may not have realised at the time that a crime was being committed. The witness can also be a victim. How a witness will react to the interview/interviewer depends on the witness's experience. Moreover, every witness is an individual who experiences events in his or her own way. This makes the interviewer's task of facilitating witness memory challenging, and as Fisher and Geiselman (1992) emphasise, the cognitive interview should not be used in the same way in all interviews: it needs to be adapted according to the witness's psychological needs and abilities. And rather than robotically using all the possible components of the CI, the most appropriate ones for a given situation should be selected. When a composite is created with a witness who has been a victim of a violent crime, describing the offender from another person's perspective might be helpful for them as it is undoubtedly difficult to keep intensely focusing on what the offender looked like. If an eyewitness had witnessed a less serious crime from further away, this mnemonic may not have the same impact.

There are five stages in the interview, which Geiselman and Fisher (2014) describe as follows: 1) An introduction stage sets the tone for the interview and establishes a relationship between the witness and the interviewer, which is regarded as highly important for a successful interview. A witness-centred approach is emphasised so that a witness feels in control of the situation and are not merely waiting for the interviewer to initiate interaction, 2) An uninterrupted free recall is then facilitated to allow the witness to explain in detail what he or she has experienced. Prior to this instruction, the witness is guided to think about the context surrounding the main event and this is the facilitation of one of the most powerful mnemonics of the cognitive interview, mental reinstatement of context (MRC). A study by Smith (1979) suggests that mental reinstatement of the learning environment may be almost as beneficial for retrieval as actual, physical reinstatement. The benefit of reinstating the physical context is easy to understand; by thinking of a daily situation where one forgot what one was doing and when s/he returned to the original place where the initial thought took place, the memory comes back. The emotional state of the individual at the time of encoding may also be recorded as part of the memory trace of the event, and lead to better recall if this state is reinstated by the witness in the interview (Bower, 1981; Tulving, 1983).

While listening to the witness, an interviewer has an opportunity to construct a strategy for eliciting additional information. Based on the contents of the uninterrupted narrative, in the next stage (3), the interviewer probes the witness on the most information-rich memory representations (scenes or mental images) and exhausts these fully to get most detail from the witness. In stage 4, the interviewer reviews the information recalled during the interview and uses further retrieval techniques, for example asking the witness to recall the information several times, in a reverse order or from another person's perspective. It is worth noting that the reverse-order technique should not be employed until after the completion of the narrative in a normal order and the follow-up questioning because although this mnemonic increases the total amount of information recalled, it was found to impair overall retrieval including reduced recall of correct information compared to free recall. As such, it should thus be used with caution (Dando et al., 2011). Lastly, at stage 5, closing of the interview takes place. This includes fulfilling official requirements (e.g., completion of a witness statement) and informing the witness that should he or she recall more information, the person can contact the interviewer afterwards. This is important information to mention due to possible delayed recollection that can occur especially following incidents that were emotionally arousing for the witness (Fisher, Brewer, & Mitchell, 2009).

About 10 years ago, it was estimated that there have been about 100 studies conducted on the cognitive interview (Fisher et al., 2010) and convincing evidence shows overall that the cognitive interview elicits more information than the other investigative interviewing methods such as the standard police interview or a structured interview. Compared to the standard interview, this effect is very robust regardless of the type of crime or witness or whether recall is immediate or delayed

(Fisher et al., 2010). For example, adults who are aged 60 years or older recalled more information when Cognitive Interview or its modified version (varied retrieval component removed) was used without increasing incorrect or confabulated statements (Wright & Holliday, 2007).

Despite unhelpful practices of the interviewers being understood for decades, cognitive interviews are still not conducted in the optimal manner, using the right techniques, thereby leaving room for improvement to enhance eyewitness memory further within the police investigations (Geiselman & Fisher, 2014). As mentioned, this has prompted other versions of the cognitive interview. A review of interview practices has revealed that the interviewers do not utilise the mental reinstatement of context component properly. In response to this, Dando et al. (2009) designed a simpler way to facilitate this powerful mnemonic by asking the witnesses to draw a detailed sketch or plan of the event they saw and describing this detail to the interviewer while drawing. While being significantly less time consuming to carry out than the usual MRC, both had a similar impact on the interview and witness's memory, making this modified MRC a viable technique. A modified MRC technique has also been used in the context of facial composite systems (Fodarella et al., 2021). This development will be discussed, in more detail, in Chapter 3. It is worth mentioning, that unlike police officers, who have many other duties in their role and therefore time constraints, facial-composite operators and particularly artists, who work on freelance basis, can conduct the cognitive interview in an unrushed manner in their specialised field. Thus, the same issues revealed in reviewing the officers' interviewing practices should not apply.

In contrast to the lack of guidelines on the technicalities of sketch composites, using the Cognitive Interview to aid retrieval of the witness's mental image of the suspect provides some standardised practice to composite construction. In England and Wales, the ACPO (Association of Chief Police Officers, 2009) guidelines recommend using the Cognitive Interview to elicit information from a witness when constructing a facial composite. It is also recommended that composite artists and operators of composite systems need to have undergone appropriate training on interviewing techniques, production of a facial image and related documentation and evidence handling and preparation. The guidelines by the International Association for Identification state that composite construction, be it a manual free hand drawing or a computerised composite, should be accompanied by an interview technique such as the cognitive interview (Richardson et al., 2010). This seems to be a somewhat looser requirement. However, a composite would be likely to raise less scrutiny in court should the artist/operator be able to answer questions about their interviewing technique (i.e., it is easier to justify forensic procedures if specific guidelines had been followed).

1.8 Enhancing recall further in facial composite construction

1.8.1 Holistic Cognitive Interview

Focusing on personality traits of a face instead of its physical features has been found to affect recognition positively (Berman & Cutler, 1998). This has also been found to apply to the effectiveness of composites, whilst encoding a face (Davies & Oldman, 1999; Shepherd et al., 1978; Wells & Hryciw, 1984). An early indication of the feature judgements being effective with sketch composites was found by Davies and Little (1990), who compared the composite sketches produced by the artist from participants’ memory and descriptions to sketches that artistically gifted undergraduate art and design students produced from their own memory. Half of the participants in the study encoded the target face under instructions to make trait judgements while the other half made ratings of facial features. Feature ratings were expected to lead to significantly better likenesses for both groups, but more marked for the experienced artist. Six colour images of targets were used in three different poses. In addition, one black and white full-face print was used. Target encoding time was three minutes which was split into 20 seconds for each colour photo and the remaining 2 minutes was spent encoding the black and white print. The police artist used no photographic reference images in her sketching technique. The composite sketches were judged for their likeness to the targets by police officers on a 7-point scale (1-no likeness at all and 7 -excellent likeness). Each sketch was shown to the judges for 10 seconds. The police artist’s composites were rated significantly better, in both trait and feature encoding conditions. Unlike hypothesised, the police artist’s composites were more recognisable likenesses when the target was encoded under a trait judgement rather than a feature judgement, which suggests that the method using no reference materials could match the holistic processing of the face better by working on the facial feature simultaneously while the witness always sees a whole face. This, however, requires a good memory to the face to be able to initiate changes into the features, and this internal process is not without challenges. Here the sketch composites were created immediately after target encoding, which is of course an unrealistic scenario.

The Holistic Cognitive Interview (H-CI) developed by Frowd et al. (2008), adds a holistic protocol to the procedure by asking the witness to reflect on the face’s characteristics for one minute and then to make character judgments (e.g., how aggressive or intelligent the person looked) on a three-point scale (low, medium and high), which aims to encourage holistic processing of the face. More recent research has demonstrated that H-CI works efficiently with composite construction, including feature systems and holistic systems (Fodarella et al., 2021; Frowd et al., 2008, 2012, 2013); for example, in Frowd et al. (2008), correct naming of PRO-fit composites constructed after the CI was 9%, increasing to 41% after the H-CI was administered. The benefits were therefore evident when features were selected in the context of a whole face, which is a normal procedure in PRO-fit composite construction.

The benefits of the H-CI have been limited to computerised composite systems (see Frowd et al., 2015) whereas pilot studies using H-CI with sketching have failed to improve composite quality markedly (although see Kuivaniemi-Smith & Frowd, 2013). Stops et al. (2012) compared a composite sketching method using CI, H-CI and H-CI sketching internal features first. The stimuli used were video clips with two characters of the TV show Eastenders involving an interaction. Cued recall was removed from the cognitive interview as field studies have indicated that doing this did not improve arrest rates when using EvoFIT (Frowd et al. 2011a), and forensically relevant detail decreases after the free recall stage, making this the most useful source for eliciting information from the witness (Roberts & Higham, 2002). Manual sketching was used to construct all composites and no reference materials (pictures of facial features in a catalogue) were used. In the CI condition, the sketch was developed as normal after the free recall stage, but no cued recall was facilitated afterwards. In the H-CI condition, the participants were asked to focus on the characteristics of the face in their mind after free recall and rate these characteristics to enhance the memory to the face in a holistic manner. In the H-CI condition, the participants were asked to focus on the internal features first (eyebrows, eyes, nose and mouth) and develop this part of the face until satisfied with the likeness, after which they were asked to focus on the external features (hair, facial shape and ears).

As hypothesised, H-CI enhanced the composite quality, with the accurate naming rate for this condition at 31% compared to 19.7% in the CI condition. Contrary to expectations, constructing the internal features first after facilitating the holistic component of the interview led to worse quality composites than those constructed in the H-CI condition, and was similar to the CI condition. This indicates that what was a hypothesised benefit of the holistic component to enhance memory for the face prior to describing it, was lost in the HCI internal first condition. Although H-CI was the best performing condition, no significant effect was found for the interview technique. To boost the statistical power of the relatively small sample size, data of CI and H-CI internal features conditions were combined (due to the similarity of their naming rates) and compared to H-CI. This revealed that H-CI produced significantly higher naming rates. Stops et al. (2012) conclude that the accuracy of the external features in the H-CI internal features condition suffers at the expense of improved quality of the internal features affecting the overall likeness of the composites. Another issue they suggest might have affected composite effectiveness is that participants were likely to spend more time on constructing the internal features first, resulting in fatigue when they focused on the external features. On reflecting why EvoFIT composites benefit from the construction of internal features first and sketches do not, Stops et al. (2012) suggest that because, with sketching, altering the features and their distances (proportions) still relies on recall in the witness-artist scenario, while the process with EvoFIT is purely based on holistic recognition, the latter method is likely to be more effective. They also state that because the witness sees only internal features for a longer period of time with no context of the whole face, this may emphasise recall further as opposed to a more holistic processing of the face. Naming of the composites has been found to reduce significantly after the holistic component was used before constructing the face on EvoFIT, when no descriptions of the face were involved, compared to CI (Frowd *et al.* 2012a). This indicates that recall and recognition complement each other and thus if these aspects are optimised for creating sketch composites, this technique could be developed further.

1.8.2 Mental reinstatement of context

Davies and Milne (1985) found that composites created in the guided memory condition were much more recognisable compared to spontaneous recall, when the construction took place in a different room than the target view. It is normal for a composite system operator to ask the witness to think back to the crime and visualise the face during the Cognitive Interview (Frowd, Nelson et al., 2012) but relatively recent research on the Mental Reinstatement of Context (MRC) mnemonic has emerged in composite construction. A revised version of the MRC (detailed CR), which requires a witness to describe the context verbally in detail, has been explored by Fodarella et al. (2021). They found some promising results on facilitating a detailed CR, when composites were constructed with PRO-fit and EvoFIT systems. The detailed CR led to composites that were significantly higher in correct naming than both physical CR and minimal CR, and EvoFIT performed significantly better than PRO-fit. A replication study found a main effect of all three factors, context (minimal and detailed), composite system (PRO-fit and EvoFIT) and interview (CI and H-CI). The composites were named significantly better in the detailed CR, H-CI and EvoFIT conditions. There was also significant interaction between CR and system: the detailed CR improved naming for EvoFIT, but not for PRO-fit. Face construction with EvoFIT was enhanced to match recent findings that indicate that composites are better likenesses when the witness focus is on the upper half of the face when it is being evolved (Fodarella et al., 2017). In the third experiment in Fodarella et al. (2021), an extended CR in which further recall on context was prompted by cued questions, was added to the context factor with minimal and detailed CR. Another factor was context attention and included incidental and intentional attention of the environment. In the latter condition, the participants were instructed to observe the environment prior to target face encoding (still image) and this, with both detailed and extensive context conditions, led to significantly better correctly named composites than composites constructed in minimal context. There was no significant difference in the different context conditions when attention was incidental. Chapter 3 will include more detail on this study.

1.9 More ecologically valid facial composite research

Koehn and Fisher (1997) suggested that future composite systems should aim to facilitate a more holistic method of processing the composite face instead of the witness selecting features that are isolated from the whole face. They point out that a face shape could be made available into which features could be added, or alternatively, the process could begin by presenting a face that already contains a complete set of features. They also propose that the operator can select these initial features based on the witness's preliminary verbal description. This method has been used with systems such as E-FIT and PRO-fit (see Fodarella et al., 2015). Both options are available for the witness, selecting features in isolation from each other and without a facial shape in view, or seeing all features at once in a whole face. The last option has indeed been found to lead to significantly more identifiable composites, compared to isolated feature selection (Skelton et al., 2015), and is the current recommended procedure when using feature systems. To further improve the methodology of facial composite research, Kovera et al. (1997) point out that while one study evaluated composites based on memory, no studies exist where both composite and target identification was based on memory. They used targets who were known to participants, students, and teachers from the same school as the participants. Their chosen target pool was based on previous research that found that recognition rates were high for former high school classmates, even after 35 years had elapsed. Only after 40 years, does recognition ability decreased markedly (Bahrick, Bahrick, & Wittlinger, 1975). This indicated to Kovera et al. (1997) that people should be generally good at recognising former schoolmates, making it a potentially useful pool for evaluating composite quality since evaluation should be carried out by people who were sufficiently familiar with the target pool. Bond and McConkey (1995) make the sensible point that a published composite is likely to be recognised by a specific pool of people who are familiar with a composite's identity, rather than by a member of the public who is unfamiliar with the face. Best practice for evaluation of composites now includes a naming task based on the above principles (see Frowd, Erickson et al., 2015).

When composites have been constructed immediately after target encoding, the same as in most earlier studies (e.g., Davies et al., 1983; Laughery et al., 1977), this design makes the studies less comparable with real life situations, than with longer retention interval. Realistically there will always be an appreciable delay between an event and composite construction, usually upwards from one day (e.g., Frowd et al., 2012b). Early research has included such a delay, for example Davies, Ellis and Shepherd (1978). Participants viewed a target face for 10 seconds and either proceeded to construct a Photofit composite immediately or returned to do so after one week. They found no significant difference between these different delay conditions when the composites were assessed by likeness ratings, sorting and identification accuracy. They concluded that, while face recall appeared to be a robust phenomenon due to no decline shown in the quality of the composites after a weeklong delay compared to those constructed immediately after target encoding, this outcome may have been affected by insensitivity of the measuring tool, Photofit. Even when the delay was three weeks, Davies et al. found no significant difference compared to composites created immediately, further highlighting the potential limitation of Photofit. Bearing in mind that research indicates that face recall does decline after a few hours from encoding (e.g., Ellis et al., 1980; Frowd & Goodfellow, 2018), which is likely to affect composite quality (Frowd & Goodfellow, 2018), this conclusion on the technical shortcomings of Photofit would appear to be credible. Contrary to Ellis et al. (1975), who found that the Photofits constructed in view were rated to have better likeness to the target than Photofits constructed from memory, Davies et al. (1978) did not replicate this effect and the composites were poorly recognised even though the conditions were ideal insofar as the target photograph was visible during the entire construction process. Target exposure time was also considered in the research. Participants saw an image of the target face for either 15 seconds or 2.5 minutes, the latter of which would be expected to produce more effective composites; however, they did not differ significantly. It was concluded that the limitations of the Photofit system was likely to be one of the major causes for this. Frowd et al. (2005b) started to address the gap in the literature by comparing different composite systems when participant-witnesses constructed composites from memory with a slightly longer, yet still short (3-4 hour) delay from the target view. Importantly, this study was a long overdue comparison of the composite systems used in the UK. Five different composite construction methods were included: E-FIT, PROfit, Sketch, Photofit and an early version of the holistic system EvoFIT. By this time, the Cognitive Interview was used in composite construction with crime witnesses (ACPO, 2009), as discussed above, and thus this study followed its guidelines. The participants viewed an image of a celebrity for one minute and constructed a composite of the face after 3-4 hours. The procedure for E-FIT and PRO-fit was essentially the same. The operator first selected features that matched the constructor's descriptions and after that showed the initial composite to them. Participants were then advised to exchange or edit the features as they wished (within PROfit) and instructed that a paint package such as Photoshop would be available for the operator to use should any features require further altering. Photofit construction followed a similar process. Participants selected features from the Photofit "Visual Index", a set of reference photographs. The operator then assembled an initial composite by slotting selected features into a template. Participants could exchange the features to better fitting ones in no particular order. Since this version of Photofit left the boundary marks on the finished composite, which risks interfering with recognition (Ellis, et al., 1978), they were removed electronically in a paint package.

Hair was not well represented in this early version of EvoFIT and therefore PRO-fit was used for participants to select a suitable hairstyle. The process started with an average-looking face being imported into PRO-fit and a hairstyle located. If hair needed altering by artistic means, it was exported into Photoshop, altered, and then imported back into PRO-fit. Participants selected six facial shapes first (from approx. 70 shape examples), then the same for colourings/textures. The face with the best overall likeness was selected as the “best-face”. These choices were then bred together, and a new screen of faces appeared. Participants had the opportunity to make changes to the “best-face” by changing the size and position of features by using a small utility within EvoFIT or to transfer the image to Photoshop and alter the facial tone. The process of selecting and breeding faces continued until the participant was satisfied that the likeness had been achieved.

The sketching method followed in this study included use of images of facial features from the FBI Facial Identification Handbook (1988), the Identikit Handbook Model II (1960) and the artist’s own selection of more recent hairstyles. This was the first stage in the composite creation, after which a light sketch was created focusing on facial proportions. Features were drawn with the help of the handbooks. While this method appears to somewhat facilitate holistic processing of the face better than the feature systems, the initial focus was on the images of features. Only then was the process allowed a shift to a more configural processing mode. Thus, this method does not appear to be the optimal sketching process, supporting more holistic face recognition, and could have contributed to worse results compared to a method that starts by creating the whole face lightly without using any reference images at this stage.

The targets were celebrities and so composite naming was used for the evaluation by asking a new group of participants to attempt to spontaneously name the composites. E-FIT and PRO-fit had the best correct naming rate (19.0% and 17.0% respectively), followed by the sketch artist (9.2%). Photofit and EvoFIT had the lowest naming rates (6.2% and 1.5% respectively). Distinctiveness was found to affect accurate naming positively in all composite systems: the more distinctive the target face, the higher the naming rate. A significant main effect was found for composite system and target distinctiveness. The interaction of these factors was also significant. E-FIT was found to be better than all other systems apart from PRO-fit, and the naming rates of PRO-fit were higher than for both EvoFIT and Photofit. In a sorting task, PROfit, E-FIT and sketch were sorted approximately with 70-80 % accuracy, and Photofit and EvoFIT were around 50%. These results largely support the naming results. For highly distinctive faces, E-FIT was better than all others except PROfit, and both performed better with distinctive faces than low distinctive faces. PROfit was equivalent to sketch, with EvoFIT and Photofit last. Sketch was best for low distinctive faces.

Frowd et al. (2005b) expected sketches to perform better in the naming task than the 9% accuracy rate found given the flexibility of the technique and the artist’s experience in the field. Since E-FIT and PRO-fit composites together were much better named than sketch, but the opposite was found with the sorting task, they concluded that this indicates that the facial features are more accurate in sketches compared to E-FIT and PRO-fit but since this is not reflected in the naming rates, they suggest that sketches lack important information for identification. An informal analysis by Frowd suggested that sketches tended to include more detail in areas such as face shape, hair, eyes, eyebrows and mouth and less detail for forehead, cheeks, chin and areas around and including the nose. He referred to similar results by Sporer (1996), with a possible exception of the nose, that these less detailed areas were also omitted in descriptions of unfamiliar faces. Frowd et al. (2005b) suggest that including more shading in the sketches could boost identification; however, this information is often not included in the witness's descriptions, and so it is difficult to see how this could be included in the sketches. It is likely that detail, such as skin texture, is difficult to recall and the witness might believe it is not an important aspect to include, being focused on more identifying features of the face.

The study design in Frowd et al. (2005b) is less comparable to the vast majority of real situations due to the short delay between the target view and the composite construction. In line with the suggestion by Kovera et al. (1997), Frowd et al. (2005a) proposed a more forensically valid delay of 2 days. This became part of a "gold" standard procedure, making the evaluation closer to the usual real-life situation. For this standard, the delay in composite construction was recommended to be 24-48 hours, faces were to be constructed by participants of unfamiliar faces, and the target pool were to be familiar to the composite evaluators, so that it was possible for the resulting composites to be named. Following these guidelines also facilitated comparison between future composite studies. Frowd et al. (2005a) evaluated the performance of E-FIT, PROfit, Sketch, FACES (a featural system used frequently in the US) and EvoFIT composite systems. The design was otherwise the same as Frowd et al. (2005b) but composites were constructed approximately 48 hours after the participant-witnesses had encoded a target face. The composites were evaluated by a naming task and a sorting task. The overall correct naming rates were again low. Sketch was found to be better than E-FIT and PRO-fit composites, and when data from both of these latter systems were combined, both EvoFIT and FACES outperformed them. In a sorting task, sketch emerged best and was significantly better than all other systems. In an identification task, in which six similar looking photos were shown for each composite, E-FIT performed significantly better than all other systems apart from sketch, and sketch was better than EvoFIT. In general, human sketching has been found to produce more identifiable composites than the feature systems when the delay from target encoding to composite construction is longer (Frowd et al., 2005a; Frowd, Erickson et al., 2015).

1.10 The role of recognition in facial composite construction

1.10.1 The development and benefits of recognition-based composite systems

The difference between internal and external feature recognition, which was discussed earlier, has been considered in the construction procedure of EvoFIT composites. One notable improvement to the process includes drawing witness's focus of attention from external features (hair and ears) to internal features (eyes, eyebrows, nose and mouth) by blurring, and thus, de-emphasising external features. This has led to improved correct naming of composites (see Frowd et al., 2008, 2010, 2011). Frowd et al. (2012) found that a very high level of blurring worked better than no blurring, low level blurring and medium level blurring. Their experiment also compared standard blur, where external feature blur was removed after face selection; extended blur, where blur was removed after use of holistic tools; and ‘infinite’ blur, essentially when internal features were only visible, and hair was selected and presented at the end. Internal features only condition was the best followed by extended blur and standard blur. So, it appears that both blurring and better yet, concealing of the external features while constructing the facial composite, leads to more identifiable composites. Later, removing the presence of external features altogether in face arrays promoted even more effective composites (e.g., Frowd et al., 2012, 2013).

Focusing on the internal features first has been found to work effectively in conjunction with the Holistic Cognitive Interview (Frowd et al., 2013). EvoFIT composites were created either by selecting the external features first which were then blurred for the face construction, or by constructing the internal features first and selecting the external features at the end of the process. Naming of the composites was best if the interview used was Holistic-Cognitive Interview rather than Cognitive Interview, the composites were viewed side-on rather than front-on, and the internal features of the composites were constructed first, then external features added (rather than having the external features blurred). All of these three techniques are effective both on their own and when combined (Frowd et al., 2013). The more effective techniques relate to holistic processing and recognition of a face (Frowd et al., 2013). Frowd et al. (2013) suggested that constructing the internal features first interferes with the processing of a face but rather than being a disadvantage, it allows the witness to focus on the internal features without having the distraction of the external features, therefore making them more identifiable. Further, more specific instructions given to witnesses to focus on the top half of the face in the initial construction stage has been found to lead to yet more identifiable composites, which is the current FBI (Fodarella et al., 2017).

The Holistic Cognitive Interview has also been found to be superior to face recall CI in Frowd, Nelson & Skelton et al. (2012), with correct naming of 39% in H-CI and 24% in CI. Feedback from the Police indicates that the H-CI works effectively for witnesses with good recall of an offender's face (Frowd et al. 2008). Field studies found that EvoFIT composites constructed in this way led to identification rate of 60% compared to 14% identification rate of E-FIT composites (Frowd et al., 2012). These techniques, combined with a post-production technique, viewing the completed composite side-on, can produce composites that are correctly named at an average as high as 74% (Frowd et al., 2013). These principles have been applied to a feature system too, where a similar level of performance is now possible (Skelton et al., 2020; see also discussion of meta-analysis in Frowd et al., 2015). Furthermore, harnessing the benefits of both the H-CI and focusing on the upper face continues to benefit the construction process. When an overall facial character judgement was facilitated first with a final focus on the target's eye region, more effective EvoFIT composites were produced (Skelton et al., 2020).

Based on research, EvoFIT is now performing better than any other system (see Frowd et al., 2011, 2013; Frowd, Erickson et al., 2015). In contrast, sketching has not been extensively researched and therefore its advantages and disadvantages are yet to be properly understood. However, sketching is clearly a flexible technique as it is not confined into templates. Therefore, it is expected that aspects of this method can be utilised in combination with feature systems such as PRO-fit and holistic systems such as EvoFIT to improve the identifiability of composites further. Currently, the author does not know of any published direct comparison of the holistic systems such as EvoFIT and sketch composites, and thus, while the former might be expected to be more effective, it remains an assumption and should be explored further. Unlike EvoFIT composites, it is not known whether a combination of post composite construction methods such as the ones used with EvoFIT (Frowd et al., 2013; Frowd, Erickson et al., 2015) would result to higher identification rates with sketches.

Frowd et al.’s studies on composite evaluation suggest that while sketching is performing worse than featural composite systems when the delay is short, it is producing better composites with a longer delay and when the memory is weaker (Frowd, 2012). In more recent studies, sketching has been found to produce somewhat more identifiable composites than the feature systems E-FIT and PRO-fit when the delay is approximately 24 hours (Frowd, Erickson et al., 2015), making the sketching technique an interesting topic to explore further.

1.10.2 Sketch artists’ use of reference materials as aid for recognition

The next section considers the different ways that sketch artists use reference materials (pictures of faces or facial features) in the sketching process. It is worth mentioning that although little empirical research exists on these techniques, anecdotal evidence is valuable since the application is practical in nature and to be an effective sketch artist requires considerable experience in both craftsmanship and interviewing. Thus, this review includes both relevant research in the area and the practical application.

In the sketch construction procedure, face recognition is involved when the witness views the initial and developing sketch. And also, when a witness views pictures of faces or facial features, for example from facial feature catalogues or a sketch artist's own collection of reference materials. This can be facilitated at different stages of the procedure, depending on the artist's practice. Some artists, however, do not use any reference materials, as they believe that seeing too many faces or facial features risks contaminating a witness's memory of the original face (e.g., Zamora, personal communication, July 14, 2017), which is in line with findings in face recognition research (e.g., Bruce & Young, 1986). However, these practitioners appear to be in the minority (Personal communication, 2011-2021). Jeanne Boylan (2000) also works largely without reference materials, particularly with traumatised victims, and criticises the standard police practice in which witnesses are bombarded with mugshots and catalogues of facial photos. Sometimes witnesses choose not to look at reference pictures during the interview, even if offered (e.g., Mancusi, 2010). As it is dependent on a witness how the composite construction proceeds, a witness cannot be forced to do so, and if they do not need any images, their mental image of the face is likely to be clear (e.g., Mancusi, 2010). In contrast, a desire not to view reference materials could indicate a weaker memory (Mancusi, 2010), and a witness is thus unable to make selections as he or she cannot clearly compare reference pictures to a mental image of the face.

Most sketch artists, on the other hand, would appear to utilise reference materials as part of their procedure (Personal communication, 2011-2021; Taylor, 2001). Two overall types of facial pictures are used by artists, whole faces and pictures showing a particular feature separately from other features (or showing only eyes and eyebrows for example). George Homa's (1983) sketching process has been described as typically including police mugshots that have been drawn from files based on the range of physiognomy displayed (Davies, 1986). Homa believes these should be limited to 6-10 photos per witness, which appears to indicate that more facial photographs could potentially interfere with the memory to a greater extent, as has been found with viewing a larger sample of mugshots (e.g., Shepherd, 1986). Taylor (2001) agrees with this notion and argues that seeing whole faces rather than individual features is better since faces do not tend to be encoded in a piece meal manner (e.g., Tanaka & Farah, 1993; Tanaka & Sengco, 1997). She emphasises the importance of achieving accuracy in the proportional representation of the face, which should be focused on drawing the initial face containing all features out of sight of the witness, and only reveal the sketch as a whole face. Taylor encourages use of reference photos after this initial stage, as she believes this will help to achieve more detailed information to the individual features without interfering with the original memory of the face. She agrees with a controlled manner of showing reference materials to witnesses and advises that they should bear relevance to the verbal description. In this way, witnesses are not required to look through hundreds of features, which is likely to keep the focus largely on the mental image of the offender and not on any newly introduced images (Taylor, 2001).

Using whole faces as a reference is a valid argument; however, it may risk contaminating witness's memory for the face being recalled. Also, law enforcement facial catalogues (e.g., FBI facial identification catalog) consist of mugshots of criminals, and there is a chance of a witness recognising an offender from these images. While seeming a useful side effect, it poses a problem to the criminal justice system as images of suspects should be presented to a witness following an official procedure. Indeed, Zamora (personal communication, 2017) describes such a situation. When he was working on a sketch composite with one witness in 1996, and the witness was looking through police mugshots of whole face reference images and recognised one of them as the offender. Zamora realised that this kind of occurrence could jeopardise an investigation and came to a decision that use of whole face reference images risks a mental image becoming distorted, possibly increasing inaccuracies in the resulting composite (cf. without use of any kind of reference materials). In the UK, facial identification guidelines advice that no photographs be shown to a witness or identification procedure be carried out before a witness has constructed a composite (ACPO, 2009). These guidelines emphasise that a witness's mental image should be protected from interference as far as possible, but they would seem to discourage use of whole face photographs as reference materials for sketch production.

Many sketch artists, for example Samantha Steinberg, advocate images of facial features that are partly blocked by a circular shape so that the whole face is not visible. Facial Identification Catalog (Steinberg, 2006, see Fig 1.9) is compiled from mugshots, and presents features in specific categories (e.g., deep set eyes, hooked noses, small lips, sunken cheeks). Unlike Taylor (2001) however, Steinberg (2006) and Mancusi (2010) instruct the witness to select facial features using the reference pictures before the initial drawing has been started. This technique exposes the witness to isolated features, which can be likened to the use of a mechanical system at this stage (see review by Shepherd & Ellis, 1996, on the mechanical systems), and which is known not to be an optimal way of processing a face when constructing a facial composite (see Skelton et al., 2015). Mancusi (2010) believes that reference materials make the face construction process easier for both a witness and an artist and help to prevent an artist misinterpreting a witness's description (as the witness's and artist's perception of a facial feature may differ). He emphasises that rather than merely copying features from reference images, the artist should aim to capture the essence of them to match the rest of the composite. This idea seems to suggest that features need to fit the context or expression of the whole face. Facial catalogues largely present faces in a neutral expression, but the witness could be describing a face having a certain expression, such as angry or laughing, and therefore this expression needs to be present in the whole face and not just for the mouth for example. It is left up to the witness to decide if a composite should present a particular expression or remain neutral, and the artist should not have influence in this decision.

In contrast to the facial identification procedure guidelines in the UK (ACPO, 2009), in the USA it is permitted that mugshots of suspects may be used as reference images in the composite construction process even if the witness has been shown them prior to the composite construction by an officer, although the artist should be aware of this as the mugshots might have affected the original memory (Mancusi, 2010). However, if a particular mugshot pointed out by the witness, is of an active suspect, the artist should not be exposed to this image due to the risk of compromising the composite sketch, which could lead to an issue in court (Mancusi, 2010).

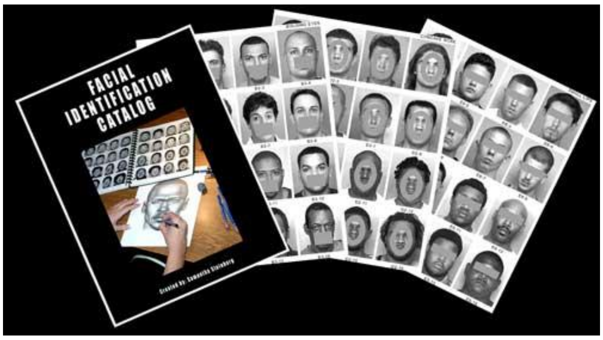


Figure 1.9 Samantha Steinberg's facial catalogue as reference images. (Steinberg, n.d.)

1.10.3 How reference materials have been used in laboratory studies

There is little evidence from early laboratory studies that focus on the sketching method and the use of reference materials. Laughery, Duval and Fowler (1977) conducted three experiments in which they employed sketch artists who were art graduates with experience in portraiture. They had practiced sketching from witness descriptions extensively and some had worked with witnesses and / or victims of real crimes constructing sketch composites. One of the variables investigated was method of construction, and sketch was compared with Identikit (the mechanical feature system used in the US). In the research, two techniques were used to obtain an initial image from a participant-witness. At first, a direct approach was defined, with the participant describing the target face guided by the artists' questions, while the artist started sketching simultaneously. The participant observed the emerging sketch and was asked to change any part of the drawing at any time. Throughout this procedure, other drawings of different faces were used as examples for comparison. This was not elaborated in the procedure, and so it is not known how exactly the other drawings were presented (i.e., were they whole faces or isolated features) and from where these drawings were drawn.

In the second approach, the participant was asked to look at a blank wall and to concentrate only on the mental image of the target. The participant described this image with the help of the artist’s guiding questions. Once the initial sketch had been completed, the drawing was shown to the witness. The witness then suggested alterations to the initial sketch. This method was suggested to interfere less with the mental image. This method is, in fact, very similar to the initial stage of the standard sketching procedure described in Fodarella et al. (2015). It seems that the direct and second approach were not compared with each other. Sketch composites were rated as having better likenesses than Identikit composites, indicating that sketching is a more flexible technique and has more potential to develop than the mechanical composites.

Frowd et al. (2005a, 2005b) included a sketch condition in their research. The participants were first instructed to look through images of facial features based on their verbal descriptions. The FBI Facial Identification Handbook (1988), the Identikit Handbook Model II (1960) and the artist's own selection of more recent hairstyles were used. A light sketch was then created focusing on facial proportions. Features were then drawn with the help of these materials. It was not explained whether the sketch was in view of the participant all the time or intermittently. Initially, pictures of separate features were seen in a catalogue, an approach that is likely to shift a witness to feature based processing. Then, the process shifts to more configural processing through first establishing proportions in the free-hand drawing with a witness, and then allowing the witness to assess the initial sketch as a complete face (Fodarella et al., 2015; Kuivaniemi-Smith et al., 2014; Taylor, 2001). It is likely that this procedure may facilitate a more natural processing of the face compared with a method that uses reference materials from the start. For this to work, though, it is likely that there needs to be a good memory of the face, which may pose a problem following a very short encounter (encoding) with a face or a long delay to face construction.

1.11 Developing sketch composites

There is almost always a delay of a varied length from the witness encoding the offender's face to the construction of a facial composite (e.g., Frowd et al., 2012b). One way to address a longer delay is to conduct the interview with the witness remotely via internet call. There has been a global shift, particularly during the Covid-19 pandemic, to increasingly move to remote online interaction (e.g., Almeida et al., 2022; Aloisi & De Stefano, 2022; Business World (India), 2020; Wild Training, 2021). This has applied to several fields including teaching, office work, gyms and personal training. The situation has led to many people working exclusively from home in those jobs where physical contact is not necessary, and thus it has become the normal way of working.

Remote interviewing in the context of composite construction is carried out by several sketch artists in the US and Australia for example due to long distances (Personal communication 2018-2021). Sketch artist Kelly Lawson from Georgia Police Department stated recently that almost all her composite interviews were conducted remotely during the Covid pandemic. She reports that it seemed to work well for her and for the witness (Lawson, personal communication, 2021). Kuivaniemi-Smith et al. (2014) formally assessed the potential of conducting the cognitive interview and composite construction remotely. One of the important benefits of this method is being able to conduct the composite interview sooner, potentially preventing memory degrading as much as it often would after many days or weeks after an incident (e.g., Ellis et al., 1980). Kuivaniemi-Smith et al. (2014) compared a remote interview to a face-to-face interview in laboratory conditions using the sketch procedure described in the next chapter. The delay from the participant witnesses seeing the target face to constructing a composite was approximately 24 hours, following the gold standard protocol (Frowd et al. 2005a). They found no significant difference between these two types of interviews. Remote interviewing for composite sketch construction has also been found to be feasible when the interviewer and interviewee are based in another country (Faundez-Salinas, 2017).

Other attempts have been made to enhance sketch composites. Robertshaw (2020) explored a novel procedure where participants were asked to recall the appearance of their target face, or not, in between encoding and creation of the sketch. The targets were characters from the Eastenders TV soap, five female and five males. The participants constructed the composites 24 hours after target encoding. A control group were interviewed and constructed the composite with a sketch artist as usual in facial composite research. Participants in the other (experimental) group did the same but were asked to write down all they remembered about the face on a piece of paper 3-4 hours after they had seen the target image. They were asked not to revisit their descriptions or bring the notes to the interview with them the next day. The correct naming rate was very good overall (M = 41.5%), similar to that found in a meta-analysis by Frowd, Erickson et al. (2015). The results of Robertshaw (2020) are promising for the sketch composites to be developed further since the composites constructed after a recall prompt were named significantly better than the control group. These results also provide more evidence that enhancing recall after an event, similar to the self-administered interview (SAI) (Gabbert et al., 2009), is beneficial to eliciting more accurate information when memory has yet to decay further (e.g., Ellis et al., 1980).

Widden et al. (2017) explored whether manual sketch composites could be improved by using photographic facial features from a digitalised facial feature catalogue and applying a Photoshop filter to reduce mid-tone information from the completed composite, which was hypothesised to remove some of the error in accuracy typically prone to facial composites. Targets used were still images of international level footballers. Two facial composites were constructed by each participant. The order of the construction varied and either the manual sketch was constructed first or a photographic composite, followed by the other construction method. A light sketch was first developed by the artist and shown to the participants only when it contained all facial features. After the participant had seen the initial sketch, he or she were able to amend it from memory. Once satisfied with the likeness, participants in the manual sketch first condition were guided to look through facial features from the digital catalogue and select the ones resembling the target face. Once the manual sketch had been finished, they proceeded to construct the photo sketch, which was done by importing facial feature pictures to Photoshop and amending them according to the participant's instructions. Participants in the photo sketch first group proceeded in the opposite order. Each finished composite was taken to Photoshop where a photocopy filter was applied. There were 80 composites constructed, eight groups of 10: manual-sketch first composites, filtered manual-sketch first composites, photo-sketch first composites, filtered photo-sketch first composites, manual-sketch second composites, filtered manual-sketch second composites, photo-sketch second composites, and filtered photo-sketch second composites. These composites were evaluated in a naming task.

The overall accurate naming rate was 14.8%, which is in line with sketches using a similar target pool (e.g., Kuivaniemi-Smith & Frowd, 2013). The filtered photo sketch composites were named significantly better than the unfiltered ones. Widden et al. (2017) Experiment 2 aimed to replicate the benefit of the best method of Experiment 1: manual sketch first, filtered photo sketch second. This revealed a significant interaction between construction (manual sketch vs. photo sketch) and presentation (original vs. simplified image), and photo sketch was named significantly better when viewed as a simplified (filtered) image. These studies provide support for an existing school of thought (example reference(s) here) that creating an initial sketch is an important step in the sketch composite procedure, and that textural information of the face contributes to the likeness of the subject.

1.12 Introduction to methodology

The methodology in all five experiments of this thesis followed the same overall design. The composites were constructed in two stages: 1) participants encoded a target face for a pre-determined duration (either a still image or a video clip), 2) participants were interviewed using a Cognitive Interview to construct a composite with the experimenter. To reflect a more realistic scenario of there being a delay in composite construction after an offender has been seen, composite construction commenced after approximately 24 hours from target encoding. The gold standard design (Frowd et al., 2005a) was followed: composite constructors were recruited to be unfamiliar with the target they see, as would be the usual case in an event of a sudden crime. Once all the composites had been constructed in one experiment, newly recruited participants evaluated the composites. One evaluation was a naming task where participants attempted to name the composites spontaneously by looking through all composites from one condition, which was randomly allocated to each participant. Naming was undertaken by participants familiar with the target pool so that the identifiability of the composite can be measured. The composites were also evaluated by a likeness rating task. For consistency across experiments, and for practical reasons (ease of recruitment), participants providing likeness ratings were recruited to be unfamiliar with the target identities. Familiar faces could also be rated differently to unfamiliar faces, so this bias was avoided. This task involved participants seeing all the composites side by side with the target image and rating the likeness of each composite to the target face on a scale of 1 (poor likeness)-7 (good likeness).

1.12.1 Standard sketch procedure

Each experiment in this thesis used the standard sketching technique, and thus the method is introduced here ahead of the experimental chapters. Other face construction methods are described in each experiment as they varied. The Cognitive Interview, which was discussed in detail earlier in this chapter, has been adopted to be used in the context of composite construction (Frowd, 2011) and its main components were used in all the experimental conditions including standard sketching. Rapport was established first to make the situation comfortable and relaxed for the mock-witnesses from the beginning. A brief introduction was then given for the composite construction process so that the face constructors knew what to expect and how long roughly it would take. It was emphasised to the participants that they should report everything they recall but to say if they could not remember something rather than to guess. Mental reinstatement of context was facilitated prior to face recall; participants were asked to think back to the situation when they saw the target face and think about the events just before and after this, how they were feeling at the time, any sounds and smells and the physical environment, the objects and people connected to it. Participants were then advised to focus on the mental image of the target face and once this was as clear as possible in their mind, to begin to recall freely, describing the target face in their own words and taking as much time as required. During this part, the researcher made written notes and listened without interruption (unless the participant spoke too quickly or too silently for information to be recorded, which is when an interviewer might ask the participant to slow down).

The researcher then proceeded to start the sketch. A faintly drawn sketch was created initially, so that it was easy to rub the pencil marks off if required at a later stage. The participants' descriptions were repeated back to them, usually proceeding from the top of the head and face downwards, and the researcher drew the features simultaneously. This often prompted the witness to offer more information about the features. The researcher also asked if the participants wish to use a pencil and paper to draw any features to support their verbal recall. More probing of detail was also often required, as part of cued recall, especially concerning proportional information. These questions were mainly open-ended such as 'can you tell me more about the width of the nose?' Multiple choice questions were also given to support recall further. This included questions such as 'can you tell me if the eyes were wide apart, close together or of average distance from each other?' Leading and suggestive questions were avoided (e.g., 'Was his nose big? 'Was the hair darker?''). The sketch was shown to the witness once all the features had been lightly drawn. From then on, participants were encouraged to guide the researcher to alter the sketch with the aim of creating the best likeness possible of the target.

The sketchpad was kept away from participants' sight while drawing, as instructed in Fodarella et al. (2015), to avoid potential interference to the mental image, unless a very minor alteration was made. The participants reviewed the sketch frequently with the aim of improving the likeness feature-by-feature. Participants were given as much time for drawing the face as required, and thus the interview proceeded at the participants' pace. They were given an opportunity to have a break and rapport was maintained throughout the interview to keep the participants motivated. The interview was concluded when participants reported that the best likeness had been achieved. See an example of a finished sketch composite in figure 1.10.

Figure 1.10 An example of a traditional sketch composite by the author.

1.13 Thesis aim and plan

Sketch artists use a sketch construction method that either relies heavily on recall and seeing the evolving sketch (standard sketching) or they use facial catalogues containing pages of facial features as reference materials to support recall for facial details. Reference materials were therefore the first variable of interest in understanding the relationship between recall and recognition in the sketching process. Thus, the impact of using reference materials was investigated in different stages of the Cognitive Interview (CI) (Geiselman & Fisher, 1985) and compared to standard sketching. Since faces are generally processed holistically (i.e., whole face context) (e.g., Tanaka & Farah, 1993; Tanaka & Sengco, 1997), it is expected that a more holistic way of seeing features will benefit the sketch composites. Target encoding time (the second variable of interest) was also considered, to model the real-world situation of a witness having only had a brief glance at the offender, and it was expected that reference materials will aid the composite quality particularly after a short encoding time as recall for detail is likely to be poor in general.

As part of attempting to develop the sketching technique, which was the third variable of interest, digital drawing and image editing possibilities were explored. Harnessing the benefits of digital methods could lead to a more fluid technique that makes editing of the composite faster and easier. Sketching on the computer (drawing on a digital drawing tablet in Photoshop or equivalent software) is already being used in the field of composites (e.g., see figure 1.11) among sketch artists (informal interviews of sketch artists by the student and LinkedIn group discussions) and there is also at least one company offering digital composites and witness interviews remotely via a video conference call (Leads Online).

Figure 1.11 An example of a digital sketch composite by the author.

The Cognitive Interview is a vital part of sketch composite construction. One of its mnemonics, Mental Reinstatement of Context (MRC), was the fourth variable of interest and was investigated with the aim of improving recall, which sketching depends on. MRC has previously been found to improve facial recognition (e.g., Malpass & Devine, 1981) and target identification (Shapiro & Penrod, 1986) and lead to more recognisable Photofit composites (Davies & Milne, 1985). More recently, the benefits of a more detailed MRC, in which the witness is asked to describe all they recall about the context, have been found to extend also to EvoFIT and PRO-fit composites (Fodarella, et al 2021). There is also evidence that intentional encoding of the environment leads to composites of better likeness compared to incidental encoding (Fodarella, Chu, Marsh et al., 2021). In addition, when the encoding was intentional, both detailed MRC and extensive (cued recall of the environmental context) MRC led to significantly better EvoFIT composites than those constructed in the minimal MRC (witness encouraged to think about the context to themselves); however, when the environmental cues were encoded incidentally there were no significant differences between the MRC conditions. The detailed and extended versions of the MRC have not yet been considered in relation to sketch composite construction and the above-mentioned results were a prompt for this technique to be investigated to find out whether it can improve the likeness of sketch composites.

The experiments are presented in separate chapters: Chapter 2, Reference materials and target encoding – Experiments 1 – 3; Chapter 3, Context reinstatement improving sketch composites – Experiments 4 and 5.

[End of chapter]

2 CHAPTER 2 – REFERENCE MATERIALS AND TARGET ENCODING

2.1 Recalling faces, its challenges, and enhancing opportunities

The context in which a face was encountered is an important cue for retrieval. For example, witnessing a face during a normal daily circumstance can have a very different outcome to witnessing or being a victim of crime, particularly if this incident occurred suddenly by a stranger and in a short space of time. If a witness had a good enough view of the perpetrator's face, it should be possible to construct a facial composite from memory, which is when recall is in a key position, especially when a sketch is constructed. The idea in featural composite systems, including sketch, is that seeing a variety of facial features will trigger recognition for a given feature and provide a point of comparison for the mental image of the face being depicted as a composite. It is widely acknowledged that recall is a difficult task and is negatively affected by delay (e.g., Davies, 1983; Ellis et al., 1980), more so than recognition (e.g., Davies, 1983), and therefore using facial images to support face recall for composite construction seems to be a valid approach. Interestingly though, while it is not surprising that composites constructed using mechanical systems have been found to be less effective than verbal descriptions due to their technical shortcomings and not matching the natural way of face processing (Christie & Ellis, 1981), superiority of verbal descriptions has been found in comparison to holistic systems too (see Lech & Johnston, 2011). In this study, the interval of time from target encoding to composite construction was three days, and EFIT-V system was expected to perform better, especially when used in conjunction with the verbal descriptions. However, the opposite result occurred. The composites were evaluated by rating a likelihood of identification (1 = very unlikely, 5 = very likely) and a matching task in which participants attempted to identify the target from an array of six faces by using the composite, verbal descriptions only, or a combination of these two. No naming task was included, which could have provided more information for the identifiability of the composites. It is possible (and suggested by the researchers) that the participants focused too much on the accuracy of featural and configural information, which is problematic in composites since they are hardly ever perfect likenesses and contain error in facial details. Only two target faces were used (Black and White targets) and composites were constructed by four White participants and four Black participants. In addition, all composite evaluators were white. There was an indication of a cross race effect, which with the small sample size could have skewed the results. To the author's knowledge, how effective hand drawn sketches are compared to verbal descriptions only, has not been formally tested. Sketching is a very interactive technique, emphasising the important communication between artist and witness, during which it is sought to clarify the meaning of the verbal descriptions more fully. Thus, it is possible that this method could supplement recall better than most composite systems.

As discussed in the Introduction chapter, archaic systems such as IdentiKit and Photofit have been found to be too rigid and with too few facial features to achieve a good resemblance (see Shepherd & Ellis, 1996). Improving these aspects in computerised systems was expected to facilitate the process of face construction, promoting more effective composites. However, these expectations have not generally been met, and recognition (naming) rates have remained low (see Frowd et al., 2015). Sketching is, by nature, a flexible system as the tool is an artist, who, if equipped with good skills, can create any shape on a piece of paper (or electronic media). The difficulty of course is retrieving sufficient detail from a witness for each individual feature, and then interpreting this information accurately. If encoding of the face has been adequate, it should be possible to retrieve an image, even if a witness initially recalls minimal information. The outcome of course depends on the interviewing skills of the artist or composite operator, which is why having experience is crucial. There is a risk of a sketch (or a composite from any system for that matter) becoming subjective if the given description is vague, or minimal in detail, as the artist cannot omit a feature in the sketch. For example, what looks like an average nose to a witness, may look different to the artist. In the standard sketch technique (see Chapter 1 for a detailed description), facial features are lightly drawn in until the whole face is represented. This face is then shown to the witness in the hope that it will trigger the recognition of the face, thus enabling the witness to guide the artist to make the sketch resemble the previously seen face as far as possible by editing features one by one.

A greater reliance on recall may also hinder recognition (Frowd, Bruce, Smith, & Hancock, 2008), via a mechanism known as the verbal overshadowing effect (e.g., Dodson, Johnson, & Schooler, 1997). The effect occurs as describing a face can create a verbal code which interferes with a code for the face that was created spontaneously during encoding (Wickham & Swift, 2006). That said, there is little evidence of the verbal overshadowing effect for face construction (Brown et al., 2020; Frowd & Fields, 2011). However, face recall is clearly important: it has been found that more identifiable composites are produced following face recall that occurs immediately after target encoding than after 24 hours (Frowd & Goodfellow, 2018). This finding resonates with a so-called Self-Administered Interview (SAI) procedure (Gabbert, Hope, & Fisher, 2009). With the SAI, witnesses recall an event, on their own, as soon as possible after a crime using a booklet containing instructions to facilitate both recall and reporting of memories. The SAI's reported avoidance of some memory loss and preservation of detail could potentially improve construction of sketch composites. More support for this finding indicating that recall can be enhanced comes from experiments by Brown et al. (2017), who found that when repeated retrieval attempts including reinstatement of context, face recall and cued recall were facilitated on the same day as when the target was encoded, and then repeated in the CI and composite construction, feature composites were constructed more recognisably. Mental reinstatement of context (MRC) has also been found to be an effective tool for enhancing recall, which is the focus of Chapter 3, and this technique to improve facial composites is considered in more detail there.

As Sporer (1996) states, descriptions of an offender are verbal reproductions of a visually perceived stimulus, while identification of a criminal (e.g., for a line-up) constitutes an act of visual recognition; construction of a composite can be classified as a visual reproduction (see Shepherd & Ellis, 1996) and be considered somewhere between these two mental processes (as described above). Constructing a sketch composite is certainly heavily recall oriented, and a witness often struggles to describe at least one feature in detail. It is in this situation that pictures of facial features may be particularly useful (e.g., Kuivaniemi-Smith et al., 2014). However, it is unclear how such reference materials may impact on the effectiveness of the sketch, or how their presence relates to encoding duration, particularly when memory is likely to be weaker after short encoding. These effects will be explored in this chapter.

2.2 Supporting the configural processing of a face

As discussed in the Introduction chapter, the theory of faces being processed in a holistic manner is backed by substantial research. Inversion of a human face disrupts the whole-face process (e.g., Yin, 1969), which leads to facial features being processed in a piecemeal manner, as is usual for other non-face objects (e.g., Young et al., 1987), and not optimal for face recognition. Configural information in the upright face's features appears to be attended to more effectively than details of the individual features themselves (e.g., Davies & Christie, 1982; Searcy & Bartlett, 1996; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). This does not mean that individual features of an unfamiliar face will not be recalled, but rather that they will be recalled better in the context of a whole face. After all, the pieces and the part where they attach to are interconnected, and thus, both configural and piecemeal mechanisms are likely to be involved (see Bruyer, 2011). Any deviation from the average proportional information of a face is emphasised by a caricature, making a face more recognisable, which also points to the importance of processing both features and their spacing (e.g., Rhodes et al., 1987; Webster et al., 2004). A dynamic caricaturing effect, in which a composite is seen moving from negative through to positive caricature states, is believed to be effective by reducing error in composites during the negative caricature states, while enhancing distinctive information in the positive caricature states (Frowd et al., 2007, 2012).

It is also worth noting, that there are situations where either-or thinking of the face processing does not always apply. For example, the own-race face effect demonstrates both configural and featural processing advantages over other-race faces (see Hayward et al., 2005). Other race faces are known to be processed in a more featural manner due to face inversion not affecting them as much as own race faces (e.g., Rhodes et al., 1989; but see Valentine, 1991). However, while the cross-race effect (own race faces are recognised better than other race faces) (see Malpass and Kravitz, 1969) has historically been focused on black and white participants in North America (e.g., Meissner & Brigham, 2001), diversifying the research to other ethnicities and geographical locations (e.g., Sangrigoli et al., 2005) indicates that the degree of interracial contact between group members affects the size of the cross-race effect (Brigham et al., 2007). In other words, the more exposed people are to other race faces, the more familiar they are with them in general. Impairment in face recognition can also lead to a different way of processing faces. For example, impairment of holistic processing has been found to be restricted to the eye region in autistic children, while other areas such as the mouth region had high significance for recognition in a whole face context (Joseph & Tanaka, 2003). The above examples suggest that there is a varying degree of holistic and featural processing mechanisms, and while the procedure of face construction for one facial composite system works for one person, it can be more difficult for someone else. We have not got answers for optimising this on an individual basis yet. Nevertheless, it would appear sensible to aim to find techniques in sketching that support holistic processing of the face due to the robust evidence of this mechanism in general for human face processing.

The internal facial features such as eyes, brows, nose, mouth, play a key role in recognising familiar faces (e.g., Ellis & Shepherd, 1992), while the external features, such as hair, are emphasised in unfamiliar face recognition (e.g., Bruce et al., 1999; Ellis, Shepherd, & Davies, 1979; Young et al., 1985). When this shift from an unfamiliar face to a familiar face occurs in the brain processing it, is unclear, but there is suggestion that even a brief encounter with an unfamiliar face leads to the internal features becoming more prominent compared to a novel face (e.g., Clutterbuck and Johnston, 2005). Configural changes to the eyes have also been found to be detected better as the face becomes more familiar through learning to remember it (O'Donnell and Bruce, 2001), which emphasises the importance of the eye region in face recognition. A witness constructs a composite of an unfamiliar face, but since it is mostly impossible to say exactly how long they were exposed to the offender's face, it cannot be measured before the interview which part of the face they will recall best. The external features (unless concealed) are likely to stay prominent however, as was the case with newly learnt faces in O'Donnell and Bruce (2001). The internal facial features require a lot of detail to be elicited from the witness, which is demanding for them, and some witnesses may struggle to find the right vocabulary. In addition, the artist often needs to ask many probing questions from the witness to be able to draw more than a template-type feature. Sometimes, this this information cannot be drawn from the witness at this stage, which risks the composite sketch becoming generic looking and lacking resemblance to the subject. Could seeing pictures of facial features help the process of creating an initial sketch?

The emphasis on holistic recognition of faces has resulted in the development of composite systems (e.g., Frowd, Hancock, & Carson, 2004; Gibson, Solomon, Maylin, & Clark, 2009; Tredoux, Nunez, Oxtoby, & Prag, 2006), the aim of which is to make face construction easier for the witness by showing whole faces (or whole-face regions) for the witness to select as opposed to detailed recall of individual features. That said, as mentioned above, modern feature systems are also capable of facilitating holistic processing to some extent by showing witnesses individual features in the context of a complete face, not just one feature at a time, a procedure that leads to more identifiable composites (Skelton, Frowd, & Speers, 2015); however, due to the poor image quality of many features in these systems, the process may benefit from an artistically skilled operator who can modify features (e.g., in Photoshop) after composite construction has been completed. Such a procedure could enhance the likeness of the composite image, adding flexibility to the system, and creating a potentially more identifiable image. This method will be used in experiments in this thesis, and PRO-fit will also be used in conjunction with sketching, as detailed below.

More generally, composite construction is likely to benefit from a combination of holistic and featural face construction principles (see Frowd et al., 2014, for featural and spatial processing of a composite) due to their interconnected nature. Thus, combining parts of the sketching procedure with a computerised method such as PRO-fit (for holistic selection of facial features) or EvoFIT (which utilises principles of holistic face recognition) may therefore result in a very effective composite system. This proposal was in fact trialled in an unpublished study by Kuivaniemi-Smith and Frowd, (Unpublished, see Appendix B). The study design was 2 x 2 between-subjects with factors of interview type (Cognitive Interview vs. Holistic-CI) and feature selection (isolated feature vs. whole face), in which participants selected features on PRO-fit after free recall (or trait judgement via H-CI). There was a significant main effect for feature selection, indicating a large benefit for whole face over isolated-feature selection. The interaction between interview and feature selection was also significant, with H-CI isolated feature selection performing much worse than H-CI whole face, while, in the CI conditions, isolated feature and whole face performed similarly. Thus, the experiment replicated the benefit of feature selection in a whole-face (cf. isolated-feature) context found by Skelton et al. (2015).

In the research, H-CI did not work with isolated feature selection at all. This finding could be explained by transfer-appropriate processing (Schooler et al., 1997). The CI tends to lead to a focus on featural information; hence it works for isolated feature selection, but when a holistic element to it is added (H-CI) it enhances holistic recognition, therefore facilitating selection by whole faces but not isolated features. No clear benefit of the H-CI whole face method was found. A replication study (Kuivaniemi-Smith & Frowd, 2013 - see Appendix B), was interested in establishing whether PRO-fit (whole face) feature selection for sketch would have a benefit over the standard sketch (i.e., where no reference materials were involved) in combination with H-CI. TV soap Coronation Street characters were used as targets. Accurate naming of composites was much higher overall, 43.8% correct (cf. 13.9% for the first study). No reliable difference between conditions was found. However, after naming data were grouped for identities in the CI condition that were constructed below and above average—that is, a medial split—the results suggest that for half of the targets that were constructed with low identification, correct composite naming markedly increased using the H-CI procedure, but decreased for the other half. This observation indicates that ease of target-face construction was related to effectiveness of the H-CI, but clearly another study to attempt to replicate the result is required.

The current research is concerned with how to enhance a witness's memory of an unidentified person and what is the best way to retrieve it as a facial composite that bears as much resemblance to the subject as possible. While it is recommended that computerised composites are constructed by selecting facial features while seeing the whole face, it remains a grey area with sketch composites what the best method is. It has not been investigated how the way the face constructor views the developing sketch and pictures of facial features, impacts the outcome of the composite sketch. Thus, the experiments in this chapter form the research question of, firstly, whether reference materials aid the retrieval of the mental image of the face being constructed and secondly, whether the support of the reference materials is particularly helpful when encoding duration is very short, just a few seconds. Thirdly, will the composite outcome differ depending on what type of reference materials are used, and at what stage of the construction process witness views them? Since no standardised guidelines on the sketching technique exist, apart from for cognitive interviewing, one of the biggest questions to start unravelling "the mystery" of sketching among the other composite systems is to ask how sketching a face can support the configural processing of a face, and the role of recognition-based techniques such as using reference materials for the witness to view while constructing the face. The archaic mechanical systems require a witness to construct the face from selecting individual features (see Laughery et al., 1977) and a sketching technique using reference materials (with individual features) at the start of the interview before the witness views the sketch can be likened to this in terms of the face processing theory (e.g., Frowd et al., 2005a, 2005b; Steinberg, 2006). Using pictures of whole faces could facilitate configural processing (e.g., Davies, 1986; Taylor, 2001); however, this technique may cause interference to the mental image of the face being recalled (e.g., ACPO, 2009; Bruce & Young, 1986). Neither of these ways of utilising reference materials is seen as optimal and so a novel way of using reference materials is tested in Experiment 2 of this chapter. Since the standard sketch method requires detailed verbalisation of facial features at the start of the process and uses no reference materials, it is investigated whether using PRO-fit for the witnesses to select features from in the whole face context can be beneficial in the procedure of otherwise sketching the face when using a Cognitive Interview. This method is expected to facilitate holistic processing of the face, particularly when encoding duration of a target is very short.

PRO-fit is the chosen computerised system in the experiments in this thesis, as it has been used in previous facial composite research regularly with ecologically-valid procedures (e.g., Fodarella et al., 2021; Frowd et al., 2005a & 2005b; Kuivaniemi-Smith & Frowd, 2013; see appendix; Skelton et al., 2015, providing a well-researched computer system with which to compare sketching. PRO-fit also uses greyscale images rather than colour images as E-FIT does, and it is thus seen to match the greyscale pencil sketches better visually. Colour images do not appear to facilitate identification of composites (e.g., Frowd et al., 2006), and using colour might add unnecessary factors of face recognition in the experiments, for example how the change from the initial colour composite being turned into a greyscale sketch would affect the mental image of the constructor. The PRO-fit system was also readily accessible and easy to use for the experiments. Using the same computerised system in all experiments was also important, for consistency.

2.3 Target encoding duration

How long the witness was exposed to an offender's face is one of the key factors in later recall and recognition of that person. The *Turnbull guidelines (ACPO, 2009*; The Crown Prosecution Service, 2018) help police officers to assess whether exposure time was adequate for the witness's testimony to be as reliable. Sometimes, a witness might say that they would be able to recognise the offender if seen again but might struggle to recall the face (Frowd, 2011). In these instances, the standard sketch technique, one that uses no recognition cues other than the developing sketch, may be too difficult for a witness to achieve a reasonable likeness to the subject. In line with expectation, relatively shorter encoding time has been found to lower recognition rates, by increasing false identification (Light et al., 1979; Memon, Hope & Bull, 2003; Shepherd, Gibling, & Ellis, 1991). A meta-analysis by Shapiro and Penrod (1986), however, found unexpectedly that in addition to longer encoding time leading to increased correct identification of the target (expected outcome), false identification also increased. It was proposed that this was possibly due to a confounding variable that has not been coded (e.g., studies that show faces relatively briefly have easier recognition tasks). Read (1995) found that when participants interacted with a target in a non-offensive situation, (i.e., 4-15 minutes as opposed to less than a minute), participants made more correct identifications in target-present line-ups, but also more incorrect identification in the target-absent line-ups. Read suggested that the false identifications might partly be due to the participants feeling that, because they saw the target for a relatively long time, they were under pressure to make a selection. This kind of misidentification after longer exposure to the person has also occurred in real cases, with misidentification occurring from several minutes to up to half an hour (Cole & Pringle, 1974; Devlin, 1976), and there is evidence that a clear view of a person does not guarantee a reliable identification afterwards (Shepherd et al., 1982). It may be expected that a longer exposure to a face would lead to increased confidence when making an identification decision; however, this may not be the case and is not a reliable indicator of the accuracy of the decision.

In Memon, Hope and Bull (2003), a group of younger (aged 17-25 year) and older (aged 59-81 years) participants saw a video of a simulated robbery during which they were exposed to the robber's face for either 45 seconds or 12 seconds. Participants in the 45 seconds conditions made more correct identifications, particularly in the target present line-up and more correct rejections in the target absence line-ups than the participants in the shorter encoding condition. In the short encoding condition, the participants' confidence was higher when they made a correct identification and lower when the identification was incorrect. In the longer encoding condition, confidence did not differ. It was concluded that increased confidence was not a reliable indicator of accuracy under long exposure, especially in a target absent line-up. The authors were unable to say whether increased accuracy in the long exposure conditions was due to increased time itself or the qualitatively superior information because factors such as pose, expression, and other aspects of appearance that had not been controlled. Similarly, in a study by Valentine, Pickering and Darling (2003) using actual line-ups in the London Metropolitan Police found that a relatively long exposure duration to a suspect's face (>1 minute) led to more identifications by witnesses than a brief exposure (<1 minute). However, they were unable to follow up on accurate identification since this would involve waiting for the court decisions on whether the suspects were guilty or not. Since they were only suspects at this stage, there might have been innocent people included. Therefore, this study reveals witness confidence level without the measure of accurate identification. If a foil was identified, however, they were able to assess the frequency of known errors, and no significant differences between exposure duration were found.

To the researcher's knowledge, no published research exists on how the witnesses' confidence in the accuracy of the composite they have constructed relates to how identifiable the face was; however, confidence ratings have been found to be positively associated with correct identification when participants were attempting to name celebrities from the composites (Frowd et al., 2012). The researcher is aware of composite practitioners asking the witness to rate the likeness of the composite at the end of the interview having used this technique herself, more for the purpose of helping the witness decide whether they are satisfied with the likeness or want to continue working on the composite. Anecdotal evidence in facial composite sketches indicate that witness confidence is not necessarily a measure of how good the memory is, and witnesses who are very confident may construct a composite that does not resemble the subject very much. The author has personally been able to evaluate this visually at the end of each experiment to determine whether how much the composite resembles the target.

As MacLin et al. (2001) emphasise, the real-world value of face recognition research investigating how a varied exposure (encoding) time influences eyewitness identification of suspects, is important. The majority of research that investigates target encoding duration (with faces as stimuli) is face recognition related (as discussed above). Constructing a composite varies with its emphasis on either recall or recognition. Frowd et al. (2015) mentioned this variable in their meta-analysis on facial composites. However, they found that only a few conditions varied from 60 seconds encoding for a still image, and thus did not conduct a formal comparison for encoding time. They did suggest, though, that encoding duration should be positively related to accurate naming, with the opposite effect for inaccurate naming, much as it is for face recognition (e.g., Shapiro & Penrod, 1986).

A more recent study by Erickson et al. (2021) manipulated encoding duration in relation to weapon focus effect. Participants saw a target face as a still image, which either included an image of a knife or not for either 10 seconds or 30 seconds.

Prior to target encoding, participants were instructed that they may be shown a picture of a threatening weapon and that they would be taking part in a two-part experiment involving a computer-based task. Encoding was found to be significant; 30 second composites were named better than 10 second ones. Weapon focus was also significant; when knife was absent, composites were named better. Interaction also revealed a significant result; presence of a knife impaired the effectiveness of the composites at 10 second encoding but not at 30 second encoding. It also revealed that 10 second encoding led to less effective composites when weapon was present but there was no significant difference for encoding duration when the weapon was absent. Since this research used a holistic composite system, the results indicate that ten seconds was an adequate duration for configural processing of the face which then matched the later retrieval of the face memory by using EvoFIT. It is possible that even shorter encoding duration, for example 5 seconds, would have had more of an impact on the face processing and, thus, retrieval of the memory, and more likely so if a featural system had been used to construct a composite. The issue of encoding duration will therefore be considered in the design of the experiments in this chapter.

2.3.1 Reference materials and encoding duration

Using, or not using, reference materials is clearly a key element of the sketching process, potentially leading to a different level of likeness of the composite, which then may affect its utility in forensic investigations. Both relying largely on facial recall and using reference materials may be problematic; the latter (i.e., use of ref materials) as it tends to draw attention to features rather than the face as a whole and the former (recall), as it is a difficult task for a witness. However, the use of these recognition aids at different stages in the interview has not been explored in detail with composite sketches, and so one aim of this project is to gain a better understanding of their overall role. The question is, are the reference materials potentially more beneficial at the start of the CI or when the sketch has been developed more—that is, when it contains more detail about the face?

Since adequate encoding of a target's face is vital for the formation of a mental image of a person, the role of reference materials will be investigated in this thesis in the context of duration of target encoding. As research literature indicates, the effects of encoding duration have received little attention for facial composites, and in studies that have manipulated it, null effects have been found. A study by Davies et al. (1978) was the first one to investigate the effectiveness of the archaic Photofit kit in a laboratory study. The stimuli were two randomly constructed Photofit faces, which the participants were instructed to reconstruct using the Photofit kit. They were shown one of the target faces for 10 seconds and had the other target face visible to them for the whole duration of the construction procedure. The reconstructed composites were evaluated by likeness ratings (on a scale of 1-7) by a new set of participants. No significant difference was found in either condition; when the target face was present or when it was constructed from memory. This is likely to demonstrate the lack of flexibility in the composite construction system used such as the case in other early studies on Photofit (e.g., Christie & Ellis, 1981; Ellis et al., 1975), and with a better performing system the difference in the conditions might be evident. Erickson et al. (2010), who investigated the weapon focus and encoding duration, also failed to find a difference by encoding (10 seconds vs. 30 seconds) for a holistic system (EvoFIT).

Since a short encoding of a person's face is likely to result in a weaker memory to the face compared to longer encoding of it, it is expected that use of reference materials would be particularly helpful in this situation. If short encoding stores a face holistically, as indicated by previous research, it is sensible to expect that viewing pictures of facial features in a whole face, rather than individually, is likely to be more beneficial for retrieval of the face. The initial sketching stage in the standard sketch method on the other hand has a holistic aspect, since the lightly-drawn face is not shown to the witness until the whole face is represented. If memory of the face is strong and the witness is able to adequately verbalise facial details to the artist, a composite may not require reference materials, as the existing sketch may itself serve as a recognition cue. If, however, memory for the face is weak, as is expected to be the case after short encoding, the initial sketch based on recall is unlikely to resemble the target, and this can potentially be corrected by the witness identifying facial features from a feature catalogue. The face processing at that stage can be thought to revert to featural processing, since the catalogue presents individual features, so that the rest of the face is always blocked out to reduce interference of seeing too many (whole) faces. Such reference materials may hinder the retrieval process; however, since it is done in combination with seeing the whole face as a sketch, it may not be as disruptive to memory as one may think.

The first experiment explored whether using reference materials supports the participants' memory in constructing a facial composite more effectively by comparing the featural composite system PRO-fit with a standard sketching method that uses no reference materials. A second factor of interest was target encoding duration; very short target encoding duration (5 sec) and longer encoding duration (30 sec) were compared, as the difference between these time intervals is likely to be sufficient to have an impact on how the face is processed at encoding. It is hypothesised that, particularly in the short encoding condition, using PRO-fit will be beneficial for the participant's retrieval process and will thus lead to composites that are more identifiable, as well as being judged to have greater resemblance to the target, than when no reference materials are used with sketching. Research suggests that faces are encoded in a configural manner when the encoding time is very short (e.g., Goffaux & Rossion, 2006). Seeing facial features in a holistic manner in PRO-fit would seem to align with this theory better than describing facial features in detail first for a sketch. Recall can be a difficult task for the witness (e.g., Davies, 1983; Ellis, 1980) and therefore the expectation was for the standard sketch condition to lead to less identifiable composites, as it relies heavily on recall. This is expected to be the case particularly in the short encoding condition, since this is likely to result in a weaker memory for the face. With a much longer target encoding time, however, the benefits of the sketching method (such as flexibility) were expected to emerge. On the other hand, the possibility to apply artistic skills to editing the initial PRO-fit images is expected to make that system more flexible, possibly increasing the likeness of the composite.

2.4 Experiment 1 – Target encoding, sketch, and feature composite system

This experiment compared two composite construction methods, one using a standard commercial composite system (PRO-fit) and one using sketch, where the face was drawn by hand. Face construction took place following either a short (5 second) or long (30 second) target encoding duration. The experiment was conducted in three stages: composite construction (Stage 1) and composite evaluation: composite naming (Stage 2) and composite likeness rating (Stage 3). The composite construction stage was in two parts: target encoding and composite construction. Participants viewed a target (in a video clip) for one of two predetermined duration and returned to be interviewed by the researcher the following day using either sketch or PRO-fit composite systems. The composite evaluation stage consisted of a naming task and a likeness rating task. Data in Stage 1 and both parts of 2 were collected from separate sets of participants.

Methodology

2.4.1 Stage 1: Composite construction

Design There were two factors, target encoding duration and method of composite construction. The short target encoding time was 5 seconds while the longer one was 30 seconds, intervals of time that were thought to be of forensic relevance and would here give rise to contrasting results. The composites were constructed either by a standard sketch method or by the PRO-fit composite system. Each participant viewed one target face and constructed one composite in a 2 x 2 between-subjects design.

Composite construction was carried out with participants who were unfamiliar with the target identities to model the real-world situation where faces are usually unfamiliar at face construction (Frowd et al., 2011). Realistically, some time usually passes before a composite is constructed after a crime, and the delay is often a day or two (Frowd, 2021), or even weeks, especially in serious crimes such as sexual assault or murder (Frowd et al., 2005). The retention interval between target encoding and composite construction was therefore chosen as approximately 24 hours (operationalised as between 20 to 28 hrs). Following these guidelines allows the current research to be consistent, facilitating comparison.In the experiment, standard sketch was used (without use of reference materials) and was compared with PRO-fit. PRO-fit has an option for facial features to be selected in a whole face context, which is the normal procedure used with this system (Frowd et al., 2005b) and was used here. While other commercial software is available that has this option (e.g., E-FIT), the benefit of selecting features in a whole face context as opposed to isolated features when constructing a composite was not demonstrated by research until Skelton et al. (2015), who found that the whole-face (cf. isolated-face) context led to more recognisable composites.

Participants

Face constructors were first- and second-year Psychology students from the University of Winchester, who received course credit for participation, and people from the community, who were offered a payment of £5. All participants were unknown to the researcher. Since the targets were Dutch celebrities, see following section, it was unlikely that participants would be familiar with these identities, the usual situation for real life crimes. There were 31 female and nine male participants with an age range from 18 to 68 (*M* = 35.8, *SD* = 16.7) years. Participants were recruited via the university's Sona research participation system, while those from the community were approach by posters and word of mouth. Ten participants were allocated into each condition.

Materials

It was important to select highly identifiable well-known people as targets, to prevent too much variation in the level of the composite evaluators' familiarity of the target face, such was the case in Frowd et al. (2005). To check this issue, a pilot study of target naming was conducted. Fifteen well-known Dutch male targets (unknown in the UK) were selected. White male targets were selected in line with most previous facial composite research that utilise targets of this demographical background (see Frowd et al., 2015), making studies more comparable. Twenty-four first year psychology students at the University of Amsterdam saw photographs of the targets, which were sequentially projected onto a screen for 30 seconds in a small classroom. To assess familiarity of the targets, the participants could either name the targets or provide descriptive, biographical information written on paper. Ten out of these 15 targets were recognised at least 85% of the time and were used as targets in this experiment. See the targets' names and occupation in Table 2.1.

1	Matthijs van Nieuwkerk	TV presenter
2	Marco Borsato	Singer
3	Paul de Leeuw	TV presenter/Singer
4	Jeroen van Koningsbrugge	Actor
5	Frans Bauer	Singer
6	Dennis Storm	TV presenter
7	Jan Smit	Singer
8	Gordon Heuckeroth	Singer/TV
9	Bram Moskowicz	Ex-lawyer/TV
10	Nick Schilder	Singer

Table 2.1 Dutch celebrities used as target individuals in Experiment 1.

Videos of these targets were used as target stimuli. They were sourced and downloaded from video content website (http://www.youtube.com) through website (http://keepvid.com) by the researchers at the University of Amsterdam (Universiteit van Amsterdam, UVA). The author was not involved in this stage, as it was important for her to remain blind to the targets until all composites had been constructed. The clips presented the targets' faces in full view and under good lighting conditions (e.g., television interviews). The targets spoke Dutch, showed a variation of facial expressions and their heads were visible in frontal and three-quarter views. For this reason, 30 seconds was deemed sufficient duration for the longer encoding, potentially providing more cues for recollection. Particularly the three-quarter view has been found beneficial for recognition of unfamiliar faces (e.g., Bruce & Young, 1998; O'Toole et al., 1998), because it offers most information about the face (Hancock, 2012).

The video editing program iMovie was used to edit the clips in the desired format and length. Each final video started and ended with 5 seconds of blank black screen in order for the researcher to be able to start the video and leave the room without being exposed to the target face. The 30 second target videos included several edited clips from the same full video so that only the target was in view, and not the interviewer for example. These transitions were edited to be smooth and hardly noticeable. The five second target video was taken from the 30 second video for consistency between conditions. After the completion of all videos, they were digitally transferred to the research team involved in this thesis. One of the researchers (not the author) confirmed that the targets were unknown in the UK.

Procedure

The factors were encoding time (5 s vs. 30 s) and composite construction method (standard sketch vs. PRO-fit). The 10 targets were constructed once per condition, and thus 40 composites were created in total. Participants viewed the target face (video) first and returned to construct the composite the following day (20 – 28 hr later). Participants were tested individually, and the CI and face construction parts of the procedure was self-paced.

Target encoding

The researcher met participants in the research laboratories at the University of Winchester and other participants from the community mainly in their own houses. Participants were given a verbal and a written briefing of face construction before they were asked to sign a consent form. The video clips containing one target in each clip were stored on the researcher's laptop and had a blank thumbnail for the researcher to remain blind to the target faces. These files were not opened by the researcher before the encoding process. The clips were labelled by a code that indicated the condition to which they belonged. Participants selected a number randomly from a list of the available numbers and the researcher located a target video with this number from a folder for target videos. Before they viewed the target, the participants were instructed to stop and close the video clip if they were familiar with the face when they first saw it. None of the faces were reported to be familiar, but if any had been, another random selection would have taken place (and the procedure repeated to locate the first face reported unfamiliar). When the participant was sitting in front of the computer screen, they were asked to put the headphones on, connected to the computer. It was thought better to use the headphones so that the researcher would not hear any sound connected to the clips to prevent her from creating any mental images of the faces based on the voice of the target. When participants were ready to view the target face, the researcher asked them to pay attention to the screen and start watching the video clip when she pressed play on the clip. The researcher started the clip to confirm that it began playing without problems. As for the first five seconds the video played a black screen, after which the target was shown, the researcher had time to leave to room for the duration of the target viewing and thus remain blind to the given face. The laboratory room was small, and the door was very close to the computer, so this was physically possible. The video clips also displayed a black screen for the last five seconds before ending with the same black screen. The researcher returned to the room after at least 45 seconds had passed, but before this she checked with the participant if the screen was black. The researcher closed the video clip and moved it to another folder so that it would not be repeated with another participant. The participant was thanked and reminded that they would need to return the next day to construct a composite of the face they had just seen. Ideally, another experimenter would have carried out the encoding, but this was not logistically possible. The described methodology was seen as a good option to keep the researcher blind to the face.

Composite construction

Each participant returned to construct a composite 20-28 hr later, or in cases where participants were from the community, the researcher returned to their homes. To model the forensic situation, a different location (room) was used than where target encoding took place, and the researcher wore different clothes than in the target view as otherwise the clothing could potentially act as a cue for face recall, an unrealistic situation. A Cognitive interview was initiated, and composites were constructed using standard sketch or PRO-fit, described below (also detailed in Fodarella et al., 2015).

The Cognitive interview (CI) is a widely used interview technique in police investigations, as a large body of research indicates that more information is elicited using the technique rather than the more standard question-and-answer interview format (e.g., Memon & Bull, 1991). The CI has been adopted to be used in the context of composite construction (Frowd, 2011) and so similarly enhances witness recall. The CI was used in all the experiments in this thesis. Rapport was established by informal discussion with the participants and the researcher first provided participants with an introduction to the interview process. It was emphasised that they should report everything but to say if participants could not remember something rather than to guess. Mental reinstatement of context was facilitated prior to face recall; participants were asked to think back to the situation when they saw the target face and think about the events just before and after this, how they were feeling at the time, any sounds and smells and the physical environment, the objects and people connected to it. Participants were then advised to focus on the mental image of the target face and once this was as clear as possible in their mind, to begin to free recall, describing the target face in their own words and taking as much time as required. During this part, the researcher made written notes and listened without interruption (unless the participant spoke too quickly or too silently for information to be recorded).

Afterwards, if the participant was assigned to sketch construction, the researcher proceeded to start the sketch by a standard sketching procedure. A faintly drawn sketch was created initially. The participants' descriptions were then repeated to them, usually proceeding from the top of the head and face downwards, drawing the individual features simultaneously. This often prompted the witness to offer more information about the features. More probing of detail was also required, as part of cued recall, especially concerning proportional information. These questions were mainly open-ended such as 'can you tell me more about the width of the nose?' Multiple choice questions are also given to support recall further. This included questions such as 'can you tell me if the eyes were wide apart, close together or of average distance from each other (it was explained that there is usually one eye's width between the eyes)?' Leading and suggestive questions were avoided (e.g., 'Was his nose big? 'Was the hair darker?''). The sketch was shown to the witness once all the features had been lightly drawn. From then on, participants were encouraged to guide the researcher to alter the sketch with the aim of creating the best likeness possible of the target.

The sketchpad was kept away from participants' sight while drawing, as instructed in Fodarella et al. (2015), to avoid potential interference to the mental image unless it was for a very minor alteration. The participants reviewed the sketch frequently with the aim of improving the likeness feature-by-feature. Participants were given as much time for drawing the face as required, and thus the interview proceeded at the participants' pace. The interview was concluded when participants reported that the best likeness had been achieved.

In the PRO-fit condition, after free recall, cued questions were asked for the participant to provide more detail of the facial features unless they had already done so. Probing was less extensive than in the sketch conditions due to participants being able to see pictures of different facial features soon after free recall which would be more to do with recognition than recall. The researcher narrowed down the facial features on PRO-fit to about 20 features per feature using the 'whole face' option before showing the composite image to the participant. From there on the participants were advised to review the initial image for its resemblance to the target and then go through each feature and select the example believed to be most like the target. It was explained to participants that if they could not find a suitable alternative, the descriptions could be changed to see more choices. However, it was pointed out that if all description options were chosen, PRO-fit may only show a couple of features, sometimes none at all, and therefore it was often helpful to enter a description of a facial feature that was more generic (less specific).

Participants were also advised that facial features could be resized and positioned on the features as required. The researcher demonstrated both in the system, and they could either choose to do this themselves or the researcher did this following their instructions. Participants were encouraged to keep selecting and altering facial features until the best possible likeness had been achieved, the same as for construction of sketches. They were also given the opportunity for the composite image to be altered by the researcher in Adobe Photoshop. For example, if a mole needed adding, this was done by drawing it on the face using a Wacom tablet and digital pen, or if the hair needed lightening or darkening this was done by using the appropriate Photoshop tools and Wacom tablet. This varied between participants, and some did not request any changes to the image, while others instructed the researcher to undertake extensive changes, especially to hair.

At the end of the face construction, participants were asked to rate the likeness of the composite based on their memory of the face on a scale of 1-7 (1 = *poor likeness* and 7 = *good likeness*). The aim of this rating was to help participants to decide whether they wanted to carry on editing the image, but in practice no one continued beyond this point. Sketches took approximately 2 hours to complete and PRO-fits took 1 hour, including the time for the interview and debriefing.

2.4.2 Stage 2: Composite naming

Composite quality was assessed by a composite naming task, which mimics the real-life scenario when a composite is circulated either internally within the police or to the public in the media. The aim is for a composite to trigger recognition with someone who is familiar to the person seeing the image, providing a lead to the investigation. The naming task therefore evaluates how identifiable are the composites. This task was completed at UVA by Psychology Masters student Marco Leijtens, a collaborator for this experiment from UVA.

Design

The DV was correct naming and collected in a 2 Encoding x 2 System between-subjects design. While participants in the composite construction stage were recruited to be unfamiliar with the targets, here, familiarity with the targets was required. The level of familiarity with the targets was assessed by showing the target photographs to the participants after they had attempted to name the composites. For composite data to be viable for analysis, the experiment specified an *a priori* rule: participants should be familiar with (correctly name) at least 70% of the targets.

Participants

Ninety-one participants, who were Dutch individuals, were recruited as an opportunity sample. There were 31 male and 60 female participants, and their age ranged from 18 to 66 (M = 36.1, SD = 16.5) years. The number of participants was based on recommendation by Simmons, Nelson and Simonsohn (2011) that when an a priori sample size cannot be computed in absence of earlier known effect sizes a minimum of 20 participants per cell is needed (Leijtens, 2016). Participants were recruited at the UVA through an online research application service and paid €5 for their participation. First year Psychology students could however choose to receive 0.5 research credits instead if wished to. 22 participants were assigned in 5 sec sketch condition, 29 in 5 sec PRO-fit, 20 in 30 sec sketch and 20 in 30 sec PRO-fit.

Materials

The participants were presented with one booklet containing composites from one system, sketch, or PRO-fit, and from one encoding delay, 5 or 30 s. In addition to composites, four “foil” images of unfamiliar males of similar age were added to the booklet to make the task more realistic (as the sought identity of the individual in the composite is not always known and to minimise guessing by these participants, (Martin et al., 2017). The composites were printed to approximately 8 cm (width) x 11 cm (height) on an A4 sheet each. The target images were printed in colour, the same as for target encoding prior to composite construction.

Procedure

Participants were tested individually, and the task was self-paced. Each participant viewed 10 composites and four foil images from one condition only, randomly selected with equal sampling, since it has been found that viewing composites from different systems could interfere with the recognition process (Frowd et al., 2005b). It was explained that the composites had been created based on other participants’ description of the memory of a picture seen briefly the previous day. Participants were also informed that the composites represented famous Dutch males, such as they might have seen on the Dutch television program “Opsporing Verzocht” (analogous to “Crimewatch” in the UK) (Leijtens, 2016). Composites plus foil images from one condition were presented to participants to name. Each person looked through the images sequentially. If participants were able to correctly name the face, or if they were able to describe unambiguously who the person was without recalling their name (e.g., the show they had appeared in and their character if they were an actor), a score of 1 was assigned. If, however, the participant was not sure of the identity, or gave a wrong name, a score of 0 was assigned. Naming of the target photographs was scored in the same way and the procedure followed to obtain naming of the target photos. Each person received a different random order of presentation for composites and target pictures.

The naming task took between 15 and 30 minutes per person, including verbal debriefing. Participants received a debriefing sheet to take away with them.

2.4.3 Stage 3: Composite likeness rating

Composite quality was assessed by a likeness rating task. Likeness ratings are usually a proxy to naming (Frowd et al., 2005b) and assess the match to the target achieved on a scale of 1-7 (1 = *poor likeness* and 7 = *good likeness*). In this experiment, the ratings took the main role in composite assessment due to the low level of correct composite naming.

Design

Likeness rating scores were the DV in a 2 Encoding x 2 System within-subjects design. Participants rated all 40 composites on a scale of 1 to 7 (1 = *poor likeness*, 7 = *good likeness*). An *a priori* rule was set for recruiting only participants who were mainly unfamiliar with the targets. This is because perceived likeness of a face varies depending on level of familiarity with the identity, with familiar (cf. unfamiliar) faces typically rated more critically, an outcome that tends to produce low ratings of likeness and worse experimental power in the ensuing analysis (Frowd, 2017). Here, participants should be familiar with (correctly name) no more than 30% of the targets. In reality, none of the Dutch celebrity targets were recognised.

Participants

A new set of participants, separate from the composite constructors, were asked to rate the likeness of the composites. Eighteen participants were recruited, students at the University of Winchester, 12 females and 6 males, aged between 18 and 35 (M = 20.3, SD = 3.8) years. They were rewarded with course credit for participation.

Materials

The composites from all conditions and their corresponding targets were cut to their printed size, 8cm (wide) x 10cm (height), and displayed so that the composites were shown side by side with the target was above.

Procedure

Participants were tested individually, and the task was self-paced. Each participant rated the likeness of all 40 composites to all ten targets on a scale of 1-7 (1 = *poor likeness*, 7 = *good likeness*), presented sequentially in a different random order for each person. Showing composites all together allows a person to be able to calibrate their rating, compare across the set, to make more accurate judgements than if ratings are done in isolation to each other. This method is comparable to a within-subjects experiment, which has more power (Jackson, 2023). Participants were first asked if the target face was familiar to them, then to rate. Participants were offered a break during the task if they wished, to limit fatigue. The task took about 20 minutes to complete. Participants were debriefed verbally and given a participant-information sheet to take away with them.

Results

2.4.4 Composite naming

This experiment explored the impact of encoding duration on the identifiability of composites from Sketch and PRO-fit systems.

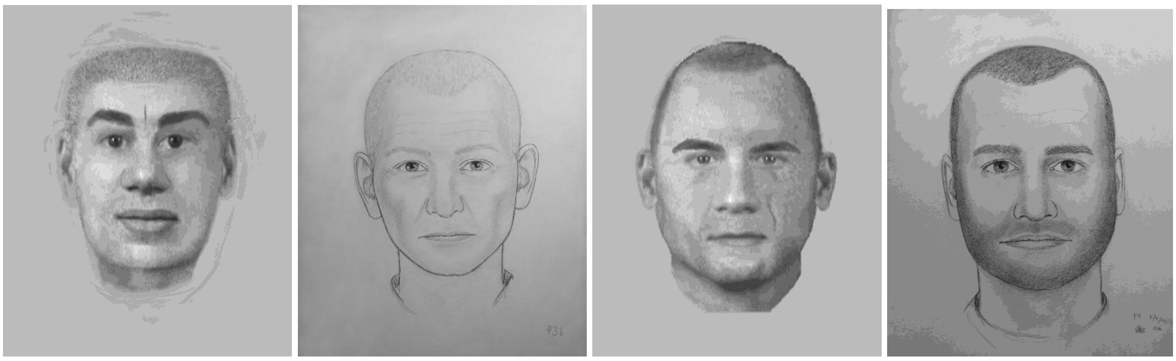
To test the hypotheses that seeing pictures of facial features improved the composites, particularly in the short encoding condition, participants attempted to name the composites and targets. Each time a participant correctly identified a composite, a score of one was assigned as described in the method section.

The target pictures were recognised with a mean of 87.7% (*SD* = 2.0%). This figure suggests that the vast majority of composites had the potential of being named correctly. As such composite naming was also considered in the context of target naming, to make the analysis fair for composite naming. For example, if the participant accurately named two composites, but only recognised eight of the targets, the naming score was calculated as two out of eight, or 25%. This measure of composite accuracy is referred to as 'conditional' naming (Frowd et al., 2005). Mean correct naming for composites using this conditional measure (Table 2.2) was very low, at 6.5% (*SD* = 16.4%) for Sketch and 3.5% (*SD* = 5.4%) for PRO-fit. As inferential analyses are unlikely to be reliable using such low values, results from the experiment were based on the analysis of likeness ratings only.

Table 2.2 Percentage Correct Conditional Naming of Composites by Encoding Duration and Construction Method, Experiment 1.

Encoding Duration (s)	*Composite System*		
	PRO-fit	Sketch	Mean
5	3.6 (5.7)	8.1 (21.4)	5.5 (14.7)
30	3.3 (5.2)	4.8 (8.1)	4.0 (6.8)
Mean	3.46 (5.4)	6.5 (16.4)	4.9 (11.8)

Note. Figures in parentheses are by-participants *SD* of the means.



5 sec PRO-fit | 5 sec sketch | 30 sec PRO-fit | 30 sec sketch

Figure 2.1 Examples of composites in all conditions with the target face from Experiment 1. The photo of the target Jeroen van Koningsbrugge (Wikimedia Commons, 2019) is licensed under the Creative Commons Attribution 3.0 Unported license. A different photo of the target was used in the experiment.

Composites in three out of the ten targets were not recognised at all in any condition, and some were recognised only once in one condition (See Table 2.3). Still, some targets received relatively much higher recognition in some of the conditions, for example Gordon Heuckeroth was recognised seven times in the five second PRO-fit condition and four times in the 30 second sketch condition.

Table 2.3 Correct Naming of Sketched Composites by items (number of times named) in Experiment 1.

Targets	5 sec sketch	5 sec PRO-fit	30 sec sketch	30 sec PRO-fit
Matthijs van Nieuwkerk	0	0	1	0
Marco Borsato	0	1	0	0
Paul de Leeuw	3	0	1	4
Jeroen van Koningsbrugge	0	1	0	0
Frans Bauer	0	0	0	0
Dennis Storm	4	2	2	1
Jan Smit	0	0	0	0
Gordon Heuckeroth	0	7	4	1
Bram Moscowicz	0	0	2	0
Nick Schilder	0	0	0	0

2.4.5 Likeness ratings

Likeness of the composites was rated on a scale of 1-7 where 1 is a *poor likeness* and 7 is a *good likeness*. Mean likeness ratings by participants are presented in Table 2.4. Overall, PRO-fit composites were rated slightly higher than Sketch composites, and composites from the longer encoding duration were rated somewhat higher than composites from shorter encoding duration.

Table 2.4 Mean Composite Likeness Ratings by System and Encoding Duration, Experiment 1.

Encoding Duration (s)	*Composite System*		
	PRO-fit	Sketch	Mean
5	3.29 (0.76)	2.71 (0.67)	3.00 (0.64)
30	3.31 (0.69)	3.74 (0.70)	3.53 (0.57)
Mean	3.30 (0.66)	3.23 (0.63)	3.26 (0.57)

Note. The Rating scale is 1 (poor likeness) .. 7 (good likeness). By-participants SD values are shown in parentheses.

By-participants analysis

A value of mean likeness rating was calculated for each participant in the individual conditions in the experiment. These mean values were analysed as two factors, for Construction Method and for Encoding Duration, using Repeated Measures Analysis of Variance. The ANOVA was significant for Encoding Duration [$F(1,17) = 27.58$, $p < .001$, $\eta_p^2 = .62$], with 30 s composites rated more accurately than 5 s composites (MD = 0.52). There was no significant effect of Construction Method [$F(1,17) = 0.27$, $p = .61$, $\eta_p^2 = .02$). The interaction between these two factors was significant [$F(1,17) = 31.42$, $p <$ 001, $\eta_p^2 = .65$]. A simple-main effect analysis revealed that while PRO-fit composites did not differ by encoding duration ($p = .87$), likeness for Sketch was higher at 30 s relative to 5 s encoding duration ($p < .001$). Also, while PRO-fit was rated significantly higher than Sketch at 5 s encoding duration ($p = .001$), the opposite was found at 30 s encoding duration, with Sketch being rated significantly higher than PRO-fit when encoding duration was 30 s ($p = .032$).

By-items analysis

To expand the understanding on the implications of the by-participant analysis on the hypotheses, the individual targets were tested by their composites in the four conditions. This by-item analysis was performed since by-participant analyses tend to be sensitive to outlier effects; composites of one target could be particularly well recognised while composites of another target might be recognised poorly if at all (Frowd et al. 2012). Therefore, conducting a by-items analysis tests whether experimental conditions generalise to other items (stimuli), in effect to check for a stimuli-as-a-fixed-effect fallacy (Clark, 1973). A value of mean likeness rating was calculated for each item (identity) in the individual conditions of the experiment. RM ANOVA of these mean values revealed that the effect of Encoding Duration was not significant [F (1,9) = 4.41, p = .07, η_p^2 = .33]. The Construction method was also not significant [F (1,9) = 0.05, p = .83, η_p^2 < .01, and neither was the interaction between these two factors [F (1,9) = 1.53 p = .25, η_p^2 < .15]. This outcome reflects the situation such that analyses by-items are usually weaker than analysis by-participants (e.g., Frowd et al., 2007); here, the benefit of the longer (cf. shorter) encoding duration only approached significance and did not interact with composite system.

Discussion

Experiment 1 explored the impact of short and long encoding duration on the resulting composites when a computerised system PRO-fit and standard sketch were used. Overall, spontaneous naming of composites was found to be very low across all conditions of the experiment. Due to this floor-level naming, no inferential statistical analysis was conducted. However, likeness ratings revealed, as hypothesised, that longer encoding led to better likeness to the target than shorter encoding. The main effect of composite system was not significant; however, it was found that 5 sec PRO-fits were rated as better likenesses than 5 sec sketches and the opposite occurred in the 30 second conditions. Also, while no significant difference was observed between 5 and 30 sec PRO-fits, 5 sec sketches were rated as much worse than 30 sec sketches.

The overall poor naming levels across the conditions suggest that the choice of the target pool could have had an impact. For composite naming, the participant-evaluators were instructed that the composites represented famous Dutch males, such as they might have seen on the Dutch television program “Opsporing Verzocht” (analogous to “Crimewatch” in the UK) (Leijtens, 2016). These famous targets were from four different occupation groups: TV presenter, singer, actor and lawyer, and some were both a TV presenter and had another occupation. Similarly low overall naming results were found by Frowd et al. (2005a, 2005b). In Frowd et al. (2005b), two main groups of celebrities were selected as targets: actors and singers in pop groups, and participants attempting to name the composites were told that they would be evaluating composites of famous people constructed in a realistic study. While targets were selected perhaps from more specific groups than this experiment did with the Dutch targets, the instruction given for naming the composites was vaguer than in this experiment. Both factors are seen as potentially problematic: target pool too large and instructions to participants attempting naming too general, and not specific enough. It is possible then that in this experiment, narrowing targets down to a couple of specific groups could have made the task more reasonable for the evaluators, possibly leading to higher naming levels overall. A large target pool may be too challenging for a facial composite study to generate reasonable naming levels, even though famous people with high familiarity were used. In this experiment, it was ensured by conducting a pilot study that the targets were generally highly familiar to participants. Independent participants rated fifteen photographs of Dutch famous male targets, and ten out of these 15 targets, who were recognised at least 85% of the time, were selected as study stimuli.

A large body of research indicates that distinctive faces are better remembered than more average looking faces (e.g., Hancock et al., 1996; Shapiro & Penrod, 1986). Whether a face is highly distinctive or not can tell us how different composite systems perform (e.g., Frowd et al. 2004, 2005a & 2005b). The level of distinctiveness was not evaluated in this experiment prior to commencing the study, but by visual inspection one might say that the targets of the composites that were not recognised by anyone (Frans Bauer, Jan Smit and Nick Schilder) look more average than those target's whose composites were recognised most often (Paul de Leeuw and Gordon Heuckeroth). It is noteworthy though that despite of the low naming rate in general, seven out of the ten targets were recognised by at least one participant in one composite condition. In real life situations, this may be enough to help the investigation by generating a lead to the identity of the offender since all viable leads should be followed up by the police.

The background of a target video could have also impacted how successfully the composite constructor encoded the target's face. In four of the target videos, audience of a talk show was visible, some very clearly. During composite construction, the author recalls some participants commenting on their focus being shifted into the audience at times. This is more likely to have occurred in the long encoding conditions, as participants had more time to look at the video. Ideally the background would not have contained other people in any of the videos, as this has likely added a distracting factor in the study that could not be controlled, and which varied between targets.

The videos also contained sound and the targets spoke Dutch. Sound as a contextual element may enhance recall of the face. However, since none of the composite constructors understood Dutch, it is unlikely they formed any meaningful associations between the words and the face. Congruent speech has been found to facilitate correct facial descriptors and lead to better likenesses of PRO-fit composites, while the opposite occurred with incongruent speech leading to poorer likenesses in the composites (Marsh et al., 2015). If the targets had been describing another person's face in English, this could have influenced the composite outcome.

The results of Experiment 1 indicate that, after 5-second encoding, a recognition-based composite construction method is likely to be more viable (a better method) than one that is based heavily on recall. This supports the finding that faces are encoded in a configural manner during very short encoding (e.g., Goffaux & Rossion, 2006). Also, because recall in composite construction is not just reliant on the memory of the witness but on their verbal capabilities too, there is likely to be a wide range of outcomes due to individual differences. In fact, describing a face is a difficult task, in which we lack expertise since we are not accustomed to doing this—due to relying more on recognition of faces in our daily lives. Some people are naturally more verbally elaborative than others and may find the composite task easier as well as being more confident in their memory, but such potential benefits do not guarantee better composites (see e.g., Wixted & Wells, 2017). In fact, the less detailed is the witness's descriptions, the more subjective sketching may become, despite the artist's best efforts to remain objective. A drawing needs to contain all facial features for normal holistic processing of the face, and therefore if the description of a nose, for example, has been reported as “average”, the artist is referring to what is average in his or her opinion if no reference pictures of noses are involved. This process thus risks an artist being more subjective. In these cases, more verbal probing, with mainly open-ended questions, is required. The verbal overshadowing effect (VOE) was coined by Schooler and Engstler-Schooler (1990), which occurs when verbally describing a face interferes with the recognition of it. Detailed facial feature descriptions could potentially interfere with the mental image of the face, but as stated earlier, evidence of the VOE for face construction is sparse (Frowd & Fields, 2011). It is reasonable to suggest though that a weaker memory due to a shorter encoding duration would be more vulnerable to a VOE. Further studies might benefit from investigating this phenomenon in the context of facial composites.

PRO-fit construction included the option of adding detail to features or editing the original PRO-fit generated image digitally in Photoshop software. This should have made the process more flexible, and it was reasonable to expect that it could potentially lead to more effective composites in terms of their likeness to the target, and thus promote more recognisable images. The results suggest that this added editing option did not improve the likeness of the composites, which was expected in longer encoding as editing is likely to have focused more on individual features, thus requiring a stronger memory. Interestingly, while 30 sec sketches were rated as being much better than 5 sec sketches, there was no marked difference between 5 and 30 sec PRO-fit composites. As mentioned in the Introduction chapter, some sketch artists provide reference materials for witnesses to look at before sketching starts (e.g., Mancusi, 2010; Steinberg, 2006). Given that the benefit of reference materials is not known empirically, this thesis intends to explore their impact with sketching at different stages of the drawing.

It is important to continue to explore the impact of shorter encoding duration on composite effectiveness, since this is when the choice of composite construction technique appears to matter more. In the current experiment, the longer encoding was 30 seconds. Target encoding used in facial composite studies is often 60 seconds, potentially giving a stronger memory than in Experiment 1, and thus further experiments here will use this longer encoding to ideally increase the impact of composites created between the shorter and longer encoding times, to explore whether such a difference may have a bigger impact on the system used—especially in the context of sketch construction (since longer encoding showed benefit). It is also of interest to examine whether reference materials could provide greater benefit to the composite process when the composite system used is overall more flexible. For this reason, the next experiment will investigate the potential benefit of providing reference pictures to participants at the start of the interview, and then continue the composite by manual sketching.

Seeing facial features together in the context of a whole face in PRO-fit was hypothesised to enhance the composite likeness. This expectation was based on considerable research indicating that selection of features in a whole-face context is more effective than seeing features in isolation (e.g., Sergent, 1986, Tanaka & Farah, 1993, 2003; Tanaka et al., 1998) Tanaka & Sengco, 1997; Valentine, 1991). Standard sketch is expected to experience the same problems as in Experiment 1, being heavily dependent on recall. It was also hypothesised that the composites created after encoding the target for 60 seconds would be closer likenesses to the targets overall and that using PRO-fit as reference materials would be helpful, especially in the short encoding condition, promoting more effective composites.

2.5 Experiment 2 – Target encoding and reference materials

This experiment followed the design of Experiment 1, except that the PRO-fit facial-composite system was now used to provide reference materials (pictures of facial features) at an early stage of composite construction, with the composite then progressed as a hand drawn sketch. This technique was compared to standard sketch. In an attempt to provide a more effective composite in the longer encoding duration, target encoding was changed from 30 to 60 seconds, with the shorter encoding duration remained at 5 seconds. Composite construction was completed a day after target encoding as before. After all composites had been constructed, they were evaluated by composite naming and likeness ratings.

Methodology

2.5.1 Stage 1: Composite construction

Design

Design in this experiment was again 2 x 2 between-subjects. The factors were encoding duration (5 seconds and 60 seconds) and composite construction method (reference materials and no reference materials). Participants in the condition using PRO-fit selected facial features first to establish the initial face in the context of the whole face and the composite was then continued as a sketch. Gold standard procedures for facial composite construction in the laboratory (Frowd et al., 2005) were followed, the same as in Experiment 1. The composites were constructed 20-28 hours after target encoding.

Participants

Face constructors were first- and second-year undergraduate Psychology students from the University of Surrey who were granted course credit for participation. Criterion for recruitment was that the participants did not follow international level football in the UK and were therefore unlikely to be familiar with the targets. There were 29 female and 11 male participants with an age range of 18-32 (M = 20.8, SD = 2.8) years. Equal number of participants were allocated to the four individual conditions of the experiment. Participants were an opportunity sample recruited via the Sona system operating within the Psychology Department.

Materials

Ten front-facing colour photographs of international level footballers were used as target faces. The criterion for the targets was that they play, or have recently played, in the premier league in the UK, and thus are well-known to those who regularly follow football. The target pool included footballers Gareth Barry, Michael Carrick, Peter Crouch, Steven Gerrard, Frank Lampard, James Milner, Scott Parker, Ryan Giggs, Andy Carroll and Robin Van Persie. All these players were prominent faces in international level football and were regularly seen on the pitch, therefore they were supposed to be recognisable for people attempting to name them from the composites in the evaluation stage. Photos were sourced on the Internet and presented a neutral expression and no distinctive features such as glasses or tattoos. The target individuals were either clean shaven or had stubble, and had a short hair (or hair tied back in a ponytail so that no longer hair was visible, like was the case with Carroll). The author was not involved in selection or setup of stimuli, to be able to remain blind to targets (so as not to be able to influence composite construction).

The images were printed in colour to dimensions of approximately 8cm (wide) x 10cm (height), each on one A4 paper.

Procedure

There were two factors, encoding time (5 seconds vs. 60 seconds) and composite construction method (standard sketch vs. sketch with reference materials). The 10 targets were constructed once per condition, meaning that 40 sketch composites were created in total. Participants were met individually by the researcher on two consecutive days, first to view a still image of a target and then to construct the composite with the researcher the following day.

Target encoding

The researcher met participants in the research laboratories at the University of Surrey and explained the procedure briefly. Occasionally, the target encoding was facilitated by one of the project supervisors, Dr Rob Nash. Participants were given a written briefing sheet to provide more information of the study after which they were asked to sign a consent form to proceed with their participation. The target images that were printed on an A4 size paper were contained in one envelope and the researcher did not see any of them until all the interviews had been completed. Participants were instructed to select one image randomly from the envelope without looking through the images while the researcher was facing away from them. Participants were asked to briefly glance at the image and, if they recognised the target, were advised to put that image back into the envelope and to select another one (this outcome did not occur). Once the participant had selected a face that was unfamiliar, the experimenter confirmed this was the case by asking the participant and instructed the participant to start looking at the image. According to allocation, participants were timed by the experimenter to look at the face for either 5 seconds or 60 seconds. Afterwards, the researcher advised the participant to place the image into an envelope for ‘used’ targets and to say when they had done so. Only after the image was inside this envelope, the experimenter faced the participant again. This was to ensure that the author did not get unblinded to the target by accident. On occasions that someone else than the author facilitated the target encoding, the blinding procedure was not necessary.

Composite construction

Participants returned to construct a composite 20-28 hours later. The image encoding and the interview were arranged to be in a different room to keep this part of the design consistent, and to model real life. Also, the same as in Experiment 1, the researcher wore different clothes for target encoding and composite construction. This design choice was followed in all further experiments too, for consistency. The Cognitive interview was conducted as described in Experiment 1, as was the standard sketch procedure. In the conditions involving reference materials, the PRO-fit composite system was used to select facial features. After free recall, the researcher first entered the participant’s description into PRO-fit to locate about 20 examples per facial feature. The participant was then asked to look at different facial features on PRO-fit to select the best features for sketching. Once all features had been selected in this way, the researcher proceeded to sketching the face as before, here using the PRO-fit face to guide the initial sketch.

At the end of the interview, all participants were asked to rate the likeness of the composite based on their mental image on a scale of 1-7 (1 = *poor likeness* and 7 = *good likeness*) to help them decide whether they wanted to continue making changes to the composite. Sketches took approximately 60 minutes and PRO-fit/sketches approximately 70 minutes to complete including the time for the interview and debriefing.

2.5.2 Stage 2: Composite naming

As in in Experiment 1, composite effectiveness was assessed by a composite naming task. Data for this was collected at the National Football Museum in Manchester by the researcher and a research assistant.

Design

In this between-subjects study, participants familiar with the targets were recruited to attempt naming the composites, and correct naming was the DV. The level of familiarity with the targets was again assessed by presenting these images to participants after they had seen the composites. The same *a priori* rule as in Experiment 1 was followed, requiring participants to be familiar with (to correctly name) at least 70% of the targets.

Participants

Twenty-seven participants were recruited as an opportunity sample, and participants were offered a small reward. There were 26 male and 1 female, aged 17-56 (M = 31.3, SD = 12.5) years. Participants were approached at the National Football Museum in Manchester and asked if they were familiar with international level footballers. Participants were assigned in equal numbers (7 participants each) to each individual conditions, except use of PRO-fit for participants encoding for 60 seconds (6 participants).

Materials

All composites created were manual hand drawn sketches (pencil/charcoal/grey scale pastels). Participant-evaluators were presented with a booklet containing composites from one condition. In addition, to make the task more realistic, four “foil” sketch images of unfamiliar males of similar age were added to each booklet. All composites were printed as before in greyscale to approximately 8 cm (width) x 11 cm (height) dimensions on a separate A4 sheets. The target images were printed in colour the same as in the target encoding task of composite construction.

Procedure

People, who were visiting the National Football Museum in Manchester, were approached and asked whether they would like to take part in a short study attempting to identify international level footballers from sketches drawn based on other participants’ description and memory of some footballers. It was emphasised that the sketches were thus not portraits. Participants were also asked if they were familiar with well-known footballers. If they met the criteria for the naming task, and agreed to participate, they were offered a small reward for participation. Participants were tested individually. First, they were given a participant information sheet to read and sign a consent form. Each participant viewed 10 composites from one condition (and the four foil sketches among these), randomly selected with equal sampling. Participants were also informed that the composites represented international level footballers, who were either current or former players in the UK Premier League. Participants were asked to bear in mind that some players might have retired or moved onto to play at a different level, because by the time the naming task was commenced, this was the case with some players. Without this additional information, the participants might have solely focused on current Premier League players and excluded other possibilities, which was thought likely to have a negative effect on overall naming. Participants looked through the composites in sequence by themselves and at their own pace, and attempted to name them. They were allowed to return to any composite they were unsure of and attempt naming again. Participants were encouraged to say a name of a footballer if one came to mind based on the composite's appearance. After the participants had finished attempting to name the composites, they were presented with the target photographs (used at the encoding stage of composite construction) in sequence and asked to name them. If they could not name the composite or the target but were able to give other (accurate) identifying information on the target individual, this was accepted as a correct answer. The order of presentation of composites and target pictures was random and different for each person. The naming task was completed in about 10 minutes per person. Participants were debriefed and given a debriefing sheet.

2.5.3 Stage 3: Composite likeness rating

As in Experiment 1, composite likeness ratings were the main method for evaluating composite effectiveness due to the floor level correct composite naming. Participants assessed the match to the target on a scale of 1-7 (1 = *poor likeness* and 7 = *good likeness*).

Design

The likeness rating scores were the DV in a within-subjects design. Participants rated all 40 composites on a scale of 1-7 (1-*poor likeness*, 7-*good likeness*) and were required to be unfamiliar with the targets.

Participants

A new set of participants separate from the composite constructors were asked to rate the likeness of the composites. Twenty-eight participants were recruited, who were a mixture of students at the University of Winchester and people from various communities, 17 females and 11 males, aged between 18 and 71 (M = 37.5, SD = 19.7) years. The former group were rewarded with course credit for participation and the latter group were volunteers. Participants were required to be unfamiliar with the footballer targets. Each participant rated all 40 composites.

Materials

The composites from all conditions and their corresponding targets were cut out to their printed size 8cm (wide) x 10cm (height) and displayed so that the composites were placed side by side with the target was above them.

Procedure

Participants were tested individually, and the task was self-paced. They were asked to rate the composites on a scale of 1-7 (1 = *poor likeness*, 7 = *good likeness*). Each participant rated the likeness of all 40 composites to all the ten targets, presented sequentially in a different random order for each person. Participants were asked to mention if the target face was familiar to them. They were offered a break during the task if required, to avoid fatigue. The task took about 20 minutes to complete. Participants were debriefed verbally and given a debriefing sheet to take away with them.

Results

2.5.4 Composite naming

Similar to Experiment 1, this experiment continued to investigate the impact of reference materials on composite effectiveness for varying time of target encoding.

The set of targets were identified very well by participants, with a mean naming rate of 95.4% (*SD* = 8.0%) correct. The same as in Experiment 1, a conditional naming score was calculated for composite naming. Overall percentage-correct means for the 5 second condition appear to be somewhat lower than means for the 60 second conditions (see Table 2.5), however, the mean correct naming was very low again, at 3.9% overall, and changed little by condition. As before, due to low values, no inferential analyses were carried out and the experimental outcome relied on results from the Likeness rating task. Table 2.6 demonstrates how often each composite was named per target.

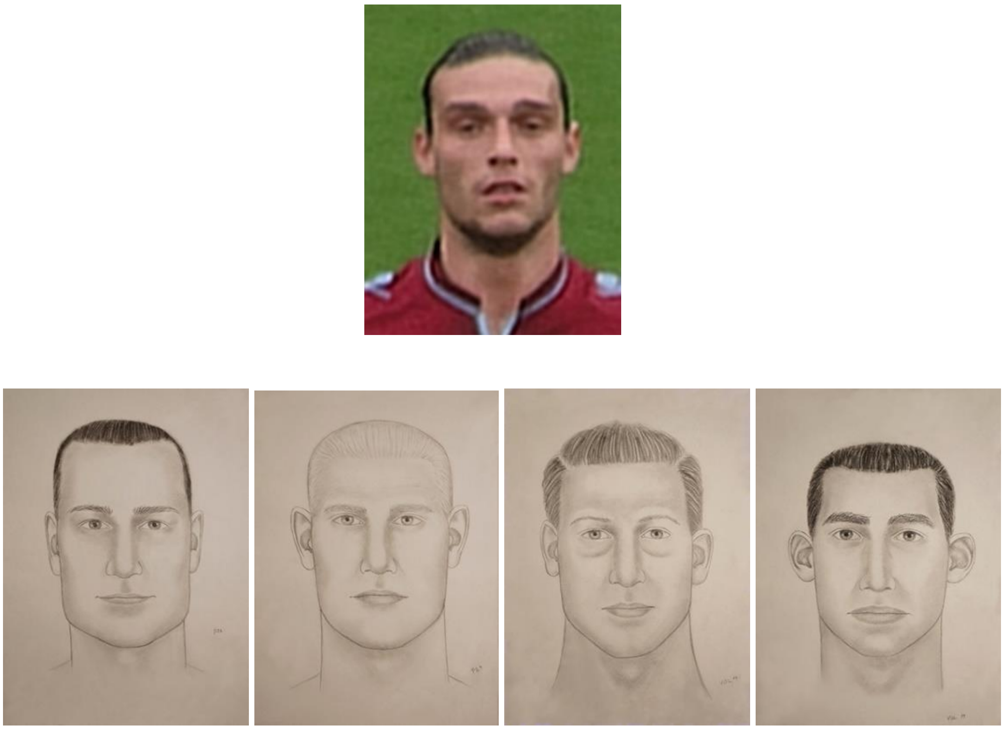
Table 2.5. Percentage Correct Conditional Naming of Sketched Composites by Encoding Duration and Construction Method, Experiment 2.

Encoding Duration (s)	*Construction Method*		
	Reference Materials	No Reference Materials	Mean
5	3.2 (5.5)	1.4 (3.8)	2.3 (4.7)
60	6.7 (8.2)	4.6 (8.6)	5.6 (8.1)
Mean	4.8 (6.8)	2.3 (4.7)	3.9 (6.6)

Note. Figures in parentheses are by-participants SD of the means.

Table 2.6 Correct Naming of Sketched Composites by items (number of times named) in Experiment 2.

<u>*Targets*</u>	5 sec sketch	60 sec sketch	5 sec PROfit/sketch	60 sec PROfit/sketch
Gareth Barry	0	0	0	0
Michael Carrick	0	0	0	0
Peter Crouch	1	0	1	1
Steven Gerrard	1	0	0	0
Frank Lambard	0	0	0	2
James Milner	0	1	0	0
Scott Parker	0	0	0	1
Andy Carroll	0	0	1	0
Ryan Giggs	0	1	0	0
Robert Van Persie	0	1	0	0



5 sec sketch   5 sec sketch/PRO-fit   60 sec sketch   60 sec sketch/PRO-fit

Figure 2.2 Examples of composites in all conditions with the target face from Experiment 2. The photo of the target Andy Carroll (Wikimedia Commons, 2012) is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. A different photo of the target was used in the experiment.

2.5.5 Likeness ratings

As in Experiment 1, participants rated the likeness of sketched composites against their corresponding target photograph. The resulting mean ratings (Table 2.7) were somewhat lower overall for composites in the 5 s conditions compared to composites in the 60 s conditions. There was little difference overall by Construction Method.

Table 2.7 Mean Composite Likeness Ratings of Sketched Composites by Encoding Duration and Construction Method, Experiment 2.

Encoding Duration (s)	*Construction Method*		
	Reference Materials	No Reference Materials	Mean
5	3.15 (0.71)	3.23 (0.81)	3.19 (0.73)
60	3.79 (0.75)	3.86 (0.73)	3.83 (0.69)
Mean	3.47 (0.68)	3.55 (0.72)	3.51 (0.68)

Note. The Rating scale is 1 (poor likeness) .. 7 (good likeness). *SD* are shown in parentheses and are by-participant values.

By-participants analysis

Conducted in the same way as Experiment 1, RM ANOVA revealed a significant effect of Encoding Duration [$F(1,27) = 68.26$, $p < .001$, $\eta_p^2 = .72$], as sketched composites were rated higher ($MD = 0.63$) when constructed from 60 s encoding than from 5 s encoding. There was no significant effect of Reference Materials

[F(1,27) = 1.66, p = .21, η_p^2 = .06] and no significant interaction between these two factors [F(1,27) = 0.02, p = .96, η_p^2 < .01].

By-items analysis

RM ANOVA revealed that the effect of Encoding Duration was marginally significant [F (1,9) = 4.52, p = .06, η_p^2 = .33], and so there was weak evidence that sketched composites were rated higher at 60 s encoding than at 5 s. The effect of reference materials was not significant [F (1,9) = 0.25, p = .63, η_p^2 = .03], and neither was the interaction between these two factors [F (1,9) < 0.01, p = .96, η_p^2 < .01].

Discussion

Experiment 2 continued to explore the impact of encoding duration on the resulting composites, comparing a relatively long and short encoding duration. PRO-fit was used to provide reference materials to support recall in the sketching process and this method was compared to standard sketch. The design was 2 x 2 with factors of encoding duration (5 and 60 seconds) and composite construction method (reference materials and no reference materials). It was hypothesised that, in the same way as in Experiment 1, composites of the longer encoding duration would be more effective than ones constructed after short encoding, and that reference materials would be beneficial at least after 5 seconds of encoding.

Similar to Experiment 1, and in spite of involving a longer encoding duration (60 seconds vs. 30 seconds), accurate naming of the composites was again very low. The target pool of international footballers was quite large, similar in principle to the celebrity target pool of Experiment 1, which is likely to have contributed to low naming. In addition to this, some of the footballers had retired by the time data collection was conducted for naming, which is also likely to have inhibited accurate naming. In the experiment, volunteers (i.e., for the naming task) were briefed that some of the footballers might not play anymore or may have moved to play football at a different level to Premier League, but these seemingly helpful instructions may not have remediated the situation. Some participants guessed composites to represent footballers who had retired long ago, for example David Beckham. This indicates that the target pool was even larger than intended by the researchers, and participants were not sure how long could have passed from a footballer's retirement from the Premier League. Since facial composites almost always contain a degree of error in the facial features and/or their configural information compared to the accurate appearance of the person being depicted, it can be a difficult task for a person to attempt identification from a composite when they are to consider all twenty teams in the Premier League. Composite evaluators were not told explicitly to consider only white footballers, and thus, while this should be mostly obvious in the composites, if there was any doubt on the ethnicity of the face depicted, the naming attempts may have included footballers from a variety of ethnicities. Six of the targets wore a football top (of their team) in the still image used for composite construction by other participants; however, even if the top was described by the composite constructor, the sketch artist (author) did not include this detail in the sketches, and so this could not be a clue for identification. Those participants, who could name the targets from their real photographs tended to know them immediately or give specific identifying detail about them, and not for example rely on the cue given by the football top. It is admitted though, that ideally, the tops could have been omitted from the images.

The chosen location (National Football Museum) for conducting naming of the composites was appropriate for finding participants who follow international level football. However, people were there at their leisure time, and some may have rather reluctantly agreed to take part, and thus, were likely to lack motivation to fully focus on the task. Some expressed frustration when they could not identify many (or any) footballers from the composites and perhaps gave up sooner than was ideal. With such composites that contain error but do in some way resemble the target, more in-depth thinking processes are required. Motivation of the participants was not measured; therefore, its impact cannot be confirmed, but of course, in ideal settings, naming would have been conducted in an unrushed manner.

While overall naming was at floor level, it is worth mentioning that all but two targets were recognised at least once in one of the composite conditions, which, as explained in the Discussion section of Experiment 1, can still be of value in real crime situation. For the benefit of the analysis of the experiment, though, the main evaluation again relied on likeness ratings. Same as in the previous experiment, composites were found to have significantly higher ratings of likeness after the relatively long than short encoding duration. In this experiment, the longer encoding duration was 60 seconds as opposed to 30 seconds, and thus, this outcome was again expected. The experiment provides more evidence for the impact of encoding duration on sketch composites, regardless of whether reference materials were utilised or not. This suggests that when the emphasis is on facial features at face construction, as it the case with sketching, having encoded the face for a longer duration is beneficial for recall of individual features.

The method of composite construction did not emerge as significant, and there was no interaction with encoding duration, and so there is evidence that reference materials (as used here) did not enhance composites overall, even in the short encoding condition as predicted. The use of PRO-fit as a method to obtain facial features was novel, and, while it had been used by researcher in a pilot study prior to conducting this experiment, the process still seemed a little clunky. Once participants had selected the initial face in PRO-fit, composite construction continued as a hand drawn sketch by the artist, copying the PRO-fit image into a sketched face. This required good skill in achieving the proportional information correctly in a situation that requires fast sketching to maintain participants' interest in the process. Naturally this phase took more time than normal (est. to be about 10 minutes more) and might have contributed to lowering motivation, inhibiting the effectiveness of the final image. It was not assessed by more technical means how successfully the researcher rendered the PRO-fit image as a sketch; therefore, this information is not available. It is possible that mistakes may have been introduced in this process ("copy errors")—although the artist (author) has considerable experience in drawing faces. This method certainly might be less than optimal if used by artists whose skills were developing. A more accurate method could potentially be trialled (e.g., one where tracing paper is placed over the computer screen for copying) to reduce subjectivity in this way of using reference materials. That said, the lack of an advantage for use of PRO-fit in this experiment would suggest that the method is not beneficial to sketch production: but then it does not appear to inhibit the effectiveness of the resulting sketch either.

The next experiment continues to investigate the use of reference materials in the context of varying encoding duration. Facial feature catalogues are commonly used by many sketch artists (e.g., Mancusi, 2010; Steinberg, 2006; Taylor, 2001), and the current researcher is trained in their use, and so this technique should be formally assessed. In the experiment, the potential accuracy of copying facial features into a hand drawn sketch was also revised using digital methods.

2.6 Experiment 3 – Target encoding and facial feature catalogue

This experiment continued to relate target encoding time to use of reference materials as a recognition aid to support recall. This time, a physical facial feature catalogue (Steinberg, 2006) was used by half of the participants to select features during composite construction. Participants not using the feature catalogue constructed composites using the standard sketch procedure. All composites were sketched manually; however, to facilitate accuracy in copying facial features, sketches were created digitally using an iPad. Stimuli were photographs of players from a local football team, Reading FC, and target encoding was either 5 seconds or 60 seconds.

Methodology

2.6.1 Stage 1: Composite construction

Design

There were two factors, target encoding duration (5 sec vs. 60 sec) and composite construction method (no reference materials vs. facial feature catalogues). The same stages were followed as previously in a 2 x 2 between-subjects design: participants viewed the targets for either 5 seconds or 60 seconds and constructed the sketches with the experimenter 20-28 hours later either using facial catalogue images as reference materials or not.

This way of using reference images from a catalogue had been practised by the artist extensively over several years, and it was hypothesised that their use would be beneficial, improving identifiability of the resulting composites. The benefit was expected to emerge particularly in the 5 second target view condition since previous experiments have found that composites are more accurate when encoding time is longer. Participants were randomly assigned, with equal sampling, to one of these four individual conditions.

Participants

Face constructors were mainly first and second year undergraduate Psychology students from Aston University, Birmingham. They were recruited via the university's Sona research participation system and received course credit for participation. A few final year students also participated voluntarily. They were recruited by Dr Robert Nash who advertised the study in one of his lectures. Participants were recruited on the basis of being unfamiliar with the targets. There were 5 males and 35 females with an age range from 18 to 30 (M = 19.9, SD = 2.4) years.

Materials

Given the low overall accurate naming in Experiment 2, in which footballer targets were used, it was seen beneficial to consider the target pool further for this experiment. It was decided that selecting players from one football team would narrow down the target pool sufficiently for composite naming level to be higher, thus allowing more useful analysis of the results. Selecting targets from one football team also allowed the researcher to collect naming data more efficiently by attending a football stadium on match days. The targets were sourced from the internet and selected by a research assistant according to predetermined criteria, after which the images were checked by the author's Director of Studies to confirm they were appropriate to be used as target stimuli. The targets were ten front-facing photographs of male footballers from Reading FC who presented a neutral expression. They were largely clean shaven. Images were printed in colour to dimensions of 8cm (wide) x 10cm (height). In the standard sketch condition, the composites were hand drawn with an iPad and apple pen using Procreate drawing program. Facial features were drawn on different layers for the ease of editing them individually, and grey scale drawing tools were used to create the sketches. In the reference materials conditions, Samantha Steinberg's facial catalogue was offered to participants to select facial features in support of their verbal descriptions.

Procedure

The factors were encoding time (5 s vs. 60 s) and composite construction method (no reference materials vs. facial feature catalogues). The 10 targets were constructed once per condition, and thus 40 composites were created in total. Participants viewed the target face (still image) first and constructed the composite the following day (20 – 28 hr later). Participants were tested individually, and the CI and face construction parts of the procedure was self-paced.

Target encoding

Another research student from Aston University assisted with target encoding. She briefed the participants about the study both verbally and by providing an information sheet to read before asking them to sign a consent form. Target encoding was then carried out as before in Experiment 2. Participants inspected the face for a specified time, as indicated by their assigned condition (either 5 or 60 seconds) in the knowledge that a composite would be created of it the following day. Participants were met in one of the study labs. Since different researchers conducted the target viewing and the composite construction most of the time, clothing was naturally different in these stages. The primary researcher conducted some of the target viewings and in these cases, care was taken (as usual) to ensure a change of clothing occurred between encoding and face construction.

Composite construction

The researcher and participants met 20 to 28 hours later to create a facial composite. As before, a Cognitive Interview was used, which started by building rapport between the researcher and the participant. An overview of the face-construction procedure was then given. After asking participants to mentally reinstate the context, free recall was facilitated. The researcher started the initial sketch on an iPad in the Procreate drawing program while simultaneously probing more verbal information about facial features, as required in the standard sketch procedure described previously. A digital Apple pen was used to draw directly on the iPad screen (12.9 inches). Procreate allows different layers to be used, which can be overlapped so that different facial features can be drawn on separate layers if it is deemed easier for later editing of the features. This option was often utilised. Once all features had been lightly established, this initial sketch was shown to participants. From there on the participants proceeded in the same was as before to develop the face, zooming in to work on each feature. When participants had reached a point when they could not think of any more changes, they were given time to review the sketch before finalising it. In conditions that involved reference materials, once the initial sketch was established, participants were then given the opportunity to peruse the facial catalogue to locate features that resembled the target's features. They were guided to look at those features that they struggled to recall, and then freely browse the catalogue. Once participants had selected one or more features, they gave the number/letter code of that feature to the researcher who noted it down and then took a photo of that feature with the iPad. The feature was then resized to anatomically fit the face and overlayed onto the sketch on a separate layer with a lowered opacity and sketched by tracing along it, yet on a separate layer, so that the original photo of the feature was preserved as guidance. This was to make the process of copying the feature into the sketch more accurate, easier, and quicker for the artist (cf. Experiment 2). Sometimes, the photo was rendered with low opacity in addition to the drawing marks, for the feature to have additional detail. This was reviewed by the witness and adjusted accordingly. The method of using reference materials varied according to how many features the participants selected. Sometimes this was just one feature, and on other occasions they selected several features. The latter naturally took longer to complete and required more revision by the participant. Standard sketches took about 75 minutes to complete and those with facial catalogue 78 minutes, including the time for the CI and debriefing.

2.6.2 Stage 2: Composite naming

The main evaluation method in this study was the composite naming task, which was completed at the Reading FC stadium in Reading, Berkshire, to recruit participants who were appropriately familiar with the target pool.

Design

The DV was correct naming and collected in a 2 (Encoding duration) x 2 (Composite construction method) between-subjects design. The level of familiarity with the targets was assessed as before, by showing the target photographs naming of the composites; as normal, participants were required to name at least 7 of the 10 targets.

Participants

Volunteers were recruited at Reading FC. They were approached in the stadium's bar where fans started to gather a few hours before the kick-off. Recruitment was also carried out outside the stadium and in the stadium's hotel. Finally, the club's head of marketing helped to directly recruit a few more participants to reach the intended sample. There were 33 male and 7 female volunteers, and their age ranged from 19 to 71 (M = 35.7, SD = 14.7) years.

Materials

Materials were the 40 greyscale composites (10 from each method of construction) and the 10 target colour photographs. Images were printed to the same dimensions as for face construction. Each set of composites included two 'foil' sketches (i.e., not of footballers from the club), created in the same way as before.

Procedure

Reading FC fans were approached at the Reading FC stadium a couple of hours before the match started and asked if they were interested in taking part in a brief study. It was explained that they would be required to attempt to identify Reading FC players from hand-drawn sketches that had been drawn from other participants' memory and description of the players whose picture they had seen a day before. The same procedure as in Experiment 2 was used to name the composites. Thus, participants were randomly assigned to one of the conditions and asked to look through one pack of ten composites and four foil images. Since one or two players had left the club by the time that naming task took place, participants were informed that the composites were of current players but due to the high turnover of the players, some may not be playing for the club anymore. The process was otherwise the same as in Experiment 2 for naming of composites and naming of target photographs.

2.6.3 Stage 3: Composite likeness rating

Since data from composite naming were sufficient for analysis, this DV was used as the main method to assess composite effectiveness in this experiment. However, composite likeness ratings also provide useful information on the effectiveness of the composites, and thus they were seen as an important inclusion to the evaluation. They also allow better comparison with Experiments 1 and 2, since these studies relied on likeness ratings. As before, participants assessed the match of sketches to the relevant targets on a scale of 1-7 (1 = *poor likeness* and 7 = *good likeness*).

Design

The likeness rating scores were the DV in a within-subjects design. Participants rated all 40 composites on a scale of 1-7 (1-poor likeness, 7-good likeness) and were required to be unfamiliar with the targets. Rather than meeting face to face, participants completed the task on Qualtrics survey website.

Participants

Participants were recruited by posting a link to a research website for psychological studies maintained by Professor John H. Krantz from Hanover College Psychology Department (https://psych.hanover.edu/research/exponnet.html), and via UCLan's Sona system. The latter group were offered course credit for completing the task. There were 36 participants, 4 males and 32 females, aged between 18 and 40 (*M* = 21.1, *SD* = 5.3) years.

Materials

Qualtrics online survey platform was used to design the likeness rating survey. Participants were instructed to use a bigger tablet or a laptop/desktop to complete the task on and not use a mobile phone due to the small screen size.

Procedure

The likeness rating task was designed in Qualtrics survey website. Each facial composite from each of the four conditions were presented with their matching target image and laid out to be viewed one after another (see Figure 2.3). Each set of images asked the participant/volunteer to rate the likeness of the composite to the target image from 1 (*poor likeness*) to 7 (*good likeness*). A box for the rating number was located under each set of images (composite and target image). Only a number code of the composite was used for the identification of each composite. Participants were also asked to write the name of the player after each set of four composites if they recognised them from the target image (which was unlikely considering targets were from a specific group of people). The survey was designed so that participants could not proceed before all questions had been answered (to avoid missing ratings).

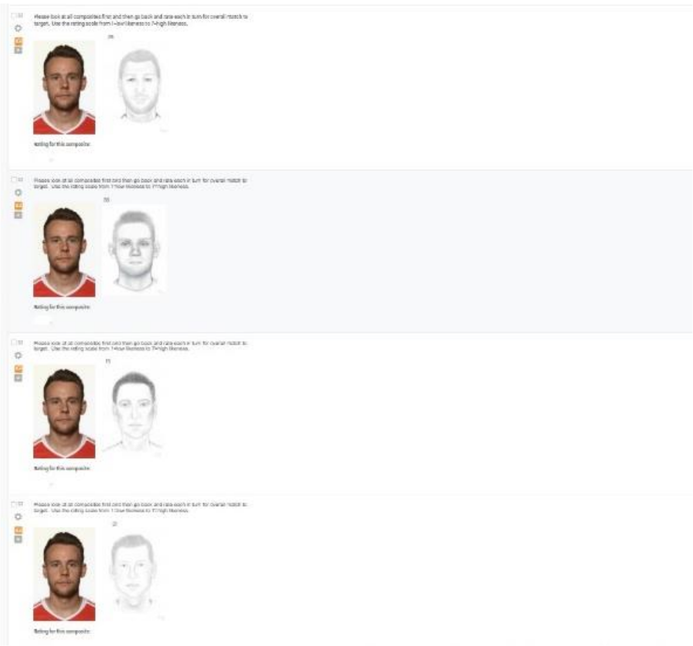


Figure 2.3 Likeness rating survey online, Experiment 3.

Results

2.6.4 Composite naming

This experiment continued to investigate the impact of reference materials on composite effectiveness relating to short and long target encoding duration. Previously, reference materials were provided by the PRO-fit system and in combination with sketching. This time, a physical copy of a facial catalogue containing depictions of facial features was used as a recognition cue for witnesses. The factors were Encoding Duration (5 s vs. 60 s) and Construction Method (no reference materials vs. facial feature catalogue). It was hypothesised that using the facial catalogues would lead to better composites, especially in the short encoding condition.

The set of targets were identified very well by participants, with a mean naming rate of 93.5% (*SD* = 0.8%) correct. The same as in Experiments 1 and 2, a conditional naming score was calculated for composite naming. Overall correct naming was 19.7% (Table 2.8), a figure that is much higher than in the previous experiments. Composites were correctly named similarly in each condition of the experiment except for 5 second without use of reference materials, where naming was much lower. Table 2.9 demonstrates the number of times the composites were named in each condition for each target.

Table 2.8 Percentage Correct Conditional Naming of Sketched Composites by Encoding Duration and Construction Method, Experiment 3.

Encoding Duration (s)	*Construction Method*		
	Facial Catalogues	No Reference Materials	Mean
5	22.1 (13.7)	7.1 (12.5)	14.1 (14.7)
60	23.7 (14.5)	26.8 (16.2)	25.2 (15.1)
Mean	22.4 (13.8)	17.0 (17.4)	19.7 (15.7)

Note: Figures in parentheses are by-participants *SD* of the means.

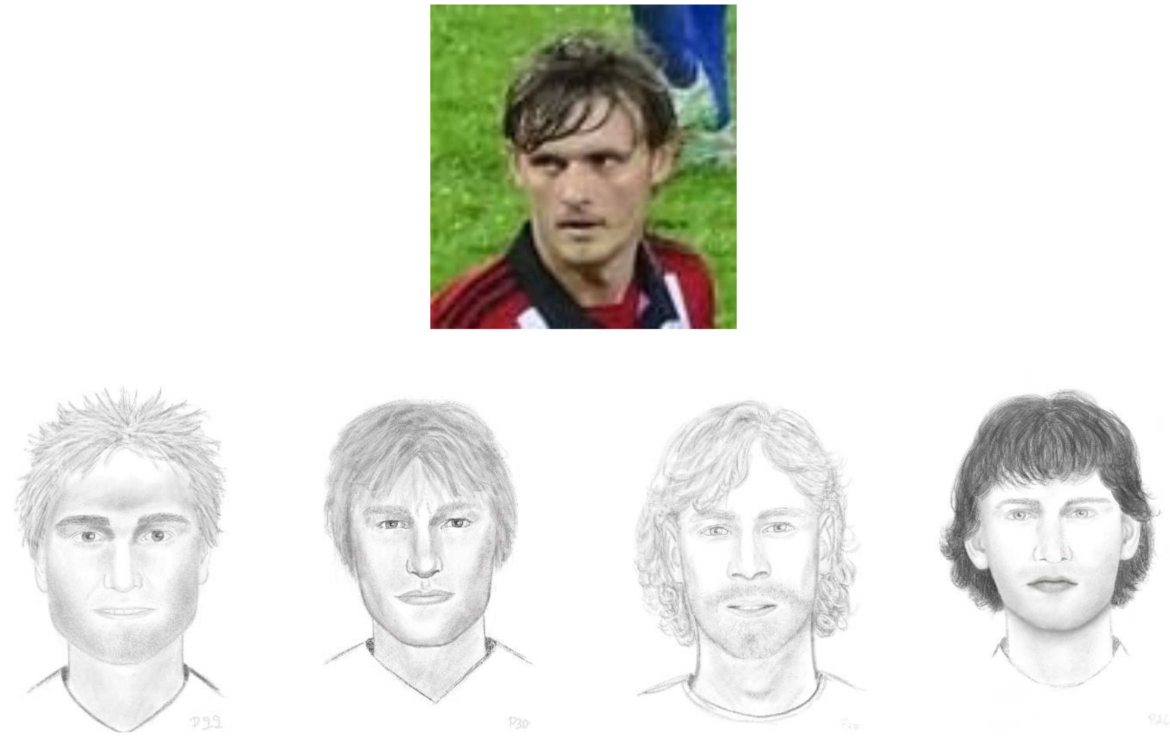
By-participant analysis

Independent Samples ANOVA revealed a significant effect of Encoding Duration [$F(1,36) = 6.05$, $p = .019$, $\eta_p^2 = .14$], as composites were named higher ($MD = 11.2\%$) when constructed from 60 second encoding than from 5 second encoding. There was no significant effect of Construction Method [F (1,36) = 1.42, $p = .24$, $\eta_p^2 = .04$]. The interaction between these two factors approached significance [$F(1,39) = 3.56$, $p = .07$, $\eta_p^2 = .09$]. A simple-main effects analysis revealed that the interaction was marginally significant as (i) the effect of encoding duration (60 s > 5 s) was only present without reference materials ($p = .004$); there was a non-significant effect of encoding duration following use of reference materials ($p = .69$), and (ii) the non-significant effect of Construction Method was restricted to the longer encoding ($p = .63$); under short encoding, reference materials led to better named composites when used than not ($p = .036$).

Table 2.9 Correct Naming of Sketched Composites by items (number of times named), Experiment 3.

Targets	5 s no Ref	5 s Ref	60 s no ref	60 s Ref
Chris Gunter	0	1	1	0
Yann Kermorgant	3	4	7	5
Roy Beerens	1	0	2	1
George Evans	0	0	3	5
John Swift	0	4	1	2
Paul McShane	0	3	0	0
Joey van den Berg	0	1	3	0
Liam Kelly	1	2	3	3
Vito Mannone	0	2	2	4
David Edwards	2	2	4	2

By-items analysis RM ANOVA revealed a significant effect of Encoding Duration [$F(1,9) = 6.08$, $p = .036$, $\eta_p^2 = .40$], with 60 second encoding producing composites that were recognised more often than 5 second encoding. There was an approaching significant effect of Construction Method [$F(1,9) = 4.26$, $p = .07$, $\eta_p^2 = .32$], as there was weak evidence that use of reference materials (cf. no reference materials) led to higher-named composites. The interaction between these two factors was not significant [$F(1,9) = 1.90$, $p = .20$, $\eta_p^2 = .17$].



5 sec no ref pictures | 5 sec ref pictures | 60 sec no ref pictures | 60 sec ref pictures

Figure 2.4 Examples of composites in all conditions with a target face from Experiment 3. The photo of the target John Swift (Wikimedia Commons, 2015) is licensed under the Creative Commons Attribution-Share Alike 2.0 Generic license. A different photo of the target was used in the experiment.

2.6.5 Likeness ratings Results

As in Experiment 1 and 2, participants rated the likeness of sketched composites against their corresponding target photograph—although here, data were collected online (cf. face to face). The resulting mean ratings (Table 2.10) were somewhat lower overall for composites in the 5 s conditions compared to the 60 s conditions. Also, composites were rated slightly higher when created using reference materials than without.

Table 2.10 Mean Composite Likeness Ratings by Construction Method and Encoding Duration, Experiment 3.

Encoding Duration (s)	Construction Method		
	Facial Catalogues	No Reference Materials	Mean
5	3.14 (0.73)	3.09 (0.69)	3.12 (0.66)
60	3.89 (0.76)	3.61 (0.79)	3.75 (0.67)
Mean	3.52 (0.65)	3.35 (0.66)	3.43 (0.62)

Note. The Rating scale is 1 (poor likeness) .. 7 (good likeness). *SD* are shown in parentheses and are by-participant values.

By-participants analysis

RM ANOVA revealed a significant effect of Encoding Duration [$F(1,35) = 58.18$, $p < .001$, $\eta_p^2 = .62$], as composites were rated higher (*MD* = 0.63) when constructed from 60 second encoding than from 5 second encoding. Construction Method was also significant [$F(1,35) = 4.53$, $p = .040$, $\eta_p^2 = .12$], as use of reference materials led to higher ratings (*MD* = 0.17). There was no significant interaction between these two factors [$F(1,35) = 2.31$, $p = .14$, $\eta_p^2 = .06$].

By-items analysis

RM ANOVA revealed that the effect of Encoding Duration was significant [$F\ (1,9) = 6.34$, $p = .033$, $\eta_p^2 = .41$], as composites were rated higher at 60 seconds than at 5 second encoding. Construction Method was not significant [$F\ (1,9) = 0.24$, $p = .64$, $\eta_p^2 = .03$] and neither was the interaction between these two factors [$F\ (1,9) = 0.39$, $p = .55$, $\eta_p^2 = .04$].

Discussion

Experiment 3 continued to explore the impact of short and long encoding duration in relation to the use of reference materials in the composite construction process. The design was 2 x 2, with factors of encoding duration (5 and 60 seconds) and composite system (no reference materials and facial feature catalogue). Based on previous findings, it was hypothesised that the longer encoding would again lead to better composites; also, that the use of reference materials would be beneficial, particularly following the short target encoding.

This experiment replicated Experiments 1 and 2, which all indicate that a longer encoding duration (30 or 60 s) leads to a stronger memory of the face (cf. 5 s), or retrieval of this memory is more compatible with a feature system under these circumstances, as measured by the resemblance of the composite to the target. In addition, composite naming in this experiment makes this argument stronger, indicating that composites are also more identifiable after longer encoding, when more time has been allowed to process individual features. These results are in line with face recognition studies (e.g., Laughery & Alexander, 1971; Memon et al., 2003; Reynolds & Pezdek, 1992), which suggests that recall and recognition are closely aligned in a composite construction process. In the by-participants analysis of the naming data, this effect of encoding duration was present when reference materials were *not* used. Also, reference materials improved composite naming at the short encoding duration. These results indicate that a weaker memory, which is likely to form after short encoding, is benefitted by a method that uses recognition (facial catalogues) to facilitate identification of features. In this experiment, reference materials were used after the initial sketch was developed. The process was therefore identical with the standard sketch condition until then. Taylor (2001) emphasises the importance of initial developing of the sketch, as the focus is not only on features but also on their spacing while processing a mental image of the face, making this stage more holistic. Once a witness has seen the initial sketch, the representation of that mental image materialises as a visual picture, which can be compared. If memory of the face is still relatively strong at this stage, the witness is able to make amendments to the face to achieve better resemblance to the subject. If, however, the mental image has already become vaguer and more unclear, witnesses are likely to depend on guessing more, affecting the resemblance negatively. A witness is also more likely to give up developing the sketch further if he or she is provided with no visual cues that can offer points of comparison, and thus, aid retrieval. As indicated by the correct naming data, viewing pictures of facial features from a catalogue (focus being on individual features), made the composites more recognisable when encoding of the target face had been short (5 seconds). When a witness sees pictures of facial features, that remind him or her of the target's face, in the composite face (drawn in by the artist), and this process appears to activate both featural and configural processing of a face (e.g., Tanaka & Sengco, 1997; Webster et al., 2004). This procedure is likely to alter the rest of the face since these two aspects influence each other and are not separate entities (see Bruyer, 2011). Using reference materials together with the sketched face that already contains all features in a whole face, perhaps works better with the sketching technique than seeing facial features immediately after free recall and some probing questions, even though features were selected in a whole face context. The latter was the case in Experiment 2.

Interestingly, likeness ratings differed from the naming results so that composites at 5 seconds did not improve reliably when reference materials were used (cf. without them). Likeness ratings can certainly be seen as more subjective than attempting to name an individual from a composite. In the latter case, a person either recognises the depicted individual or not, but rating a likeness of a composite while the target face is visible can introduce several factors in the evaluation. The evaluators were instructed to judge the overall likeness of the face, bearing in mind that the composite image is not a portrait but a visual representation of someone else's memory of the face, which he or she was unfamiliar with while encoding it. It is out of the experimenters' control how exactly an evaluator processes composite faces, and which features he or she thinks of as more salient, and this varies between participants. In the experiments in this chapter, participants were asked to rate the likeness of the composites while all composites from each condition were in view and the corresponding target image were visible at the same time. This method was seen to allow participants to make more accurate judgements, and likened to a within-subjects experiment, which has more power (Jackson, 2023). In other studies (e.g., Bruce et al., 2002; Fodarella et al., 2021), composites have been presented one composite at the time alongside its target face. Fodarella et al. (2021) mention that, in their experiments in general, likeness ratings corresponded with the correct composite naming results. This could be an indication that presenting composites individually with the target face could provide more reliable results. In the experiments 1 and 2 in this chapter, likeness ratings were the main evaluation method due to low overall correct naming levels, and thus, it cannot be said whether the two evaluation methods were in line with each other. It is worth mentioning that likeness ratings were collected via an online survey rather than face to face with participants. There could be a difference between these methods, but this was not compared in this experiment. The online method is less controlled; however, participants might be more critical with their ratings as the experimenter is not in the same physical space, which could make them not to be as harsh with their judgements. This is likely to affect all conditions in the same way and is therefore not seen as problematic.

Of course, it is good to be reminded, that while naming is seen as the primary evaluation method in facial composite studies (see Frowd et al. 2015), it is not void of factors that might influence the naming results. One of these factors may be a relatively small target pool, which could lead to guessing more identities than would occur with a larger target pool (cf. international level footballers). While a potential for this element is present in all study conditions, it cannot be excluded that one condition had more participants who generated more guesses, which could have affected the results. This is more likely to have a smaller impact on the overall results, however. The overall correct naming in this experiment was much higher than in Experiments 1 and 2. This is likely to be, at least in part, due to the different target pool sizes. In this experiment, the targets were selected from one football team only instead of multiple teams. Participants' level of familiarity with the targets was, however, similar in Experiment 2 (international level footballers from multiple teams) and Experiment 3 (players from one League 1 football team), and this is likely to be because participants were approached in a place where people would have a special interest in football. Experiment 2 target pool was wider than Experiment 3, as there were many more teams and players to consider. This experiment allowed the focus to be in a smaller group, which naturally eliminates many other players as potential identities to name.

Summary of the experiments

To summarise Experiments 1, 2 and 3, it was investigated whether the use of reference materials would support recall, and how this relates to encoding duration, particularly when memory is likely to be weaker after a brief encoding. In Experiment 1, the encoding duration was 5 seconds versus 30 seconds, and standard sketch was compared to PRO-fit, which presents photographic facial features in a whole-face context. Due to low composite naming rate in general, evaluation was based on likeness ratings, which revealed, as hypothesised, that longer encoding led to better likeness to the target than shorter encoding. No main effect was found for composite system; however, 5 second PRO-fits were rated as better likenesses than 5 second sketches, and the opposite occurred in the 30 second conditions. Also, while no significant difference was observed between 5- and 30-seconds PRO-fits, 5 second sketches were rated as much worse than 30 second sketches. Experiment 2 compared 5 second encoding duration to 60 second encoding, and sketching was used in all conditions, but this time PRO-fit was used to provide reference materials for participants to select before a composite was continued as a sketch. Similar to Experiment 1, naming was very low, and longer encoding led to better likenesses than shorter encoding, but no difference was found between systems and there was no interaction. Experiment 3 used the same encoding durations as Experiment 2, and standard sketch was compared to a sketching method that used physical facial feature catalogues as reference materials. Again, longer encoding led to superior composites, both by composite naming and likeness ratings. No main effect of system was found; however, in the 5 second standard sketch, naming was much lower compared to other conditions and the interaction approached significance. This was due to the effect of encoding duration (60 s > 5 s) being only present without reference materials; there was a non-significant effect of encoding duration following use of reference materials. Also, the non-significant effect of Construction Method was restricted to the longer encoding; under short encoding, reference materials led to higher named composites when used than not. Likeness ratings revealed that 60 second sketches were rated significantly higher than 5 second composites. Composite system was also significant and using reference materials led to better likenesses.

Given the results of the first three experiments, it is reasonable to conclude that longer encoding of a face leads to better sketched composites, especially if the technique relies on recall and the developing sketch. After short encoding, supporting a witness's recall by allowing eyewitnesses to see pictures of facial features, appears to be important, but depends on how the reference materials are presented and at which stage of composite construction. The benefit of reference materials does not appear to emerge with longer encoding in sketch construction. This is suggested to be because a witness has a stronger memory and is thus able to describe facial features (and their spacing) in more detail, and so there is less need for support from reference materials as cues for recognition. Twenty-five percent overall correct naming after 60 second encoding is still far from ideal though, which indicates that the use of reference materials was not optimised and there may be a better way to get more benefit from them. Alternatively, techniques to improve recall, coupled with the use of facial feature catalogues as reference materials, could improve correct naming further. Enhancing recall is what the next experimental chapter will investigate.

3 CHAPTER 3 – CONTEXT REINSTATEMENT IMPROVING SKETCH COMPOSITES

Perceiving and processing sensory information to see, hear, taste, or feel objects in the world guides what actions are taken with respect to these objects (Sekuler & Blake, 2002). This is a foundation for any memory to form and someone to be recognised after seeing them. Visual perception is a product of bottom-up and top-down processes (Smith & Kosslyn, 2014): Bottom-up processes are driven by sensory information from the physical world and top-down processes are driven by our knowledge, beliefs, expectations and goals. Both of these processes are almost always present in any kind of perception (Smith & Kosslyn, 2014). It has been acknowledged that the problem with recalling information from our episodic memory originates from the retrieval process and the inaccessibility of information, also through interference, rather than from its capacity or how information is stored (Surprenant & Neath, 2009; Tulving & Pearlstone, 1966; also see Discussion in Capaldi & Neath, 1995). The memory trace is composed of several features (Bower, 1967; Wickens, 1970) and the effectiveness of a retrieval cue depends on the amount of feature overlap with the encoded event (Flexser & Tulving, 1978). A phenomenon called Encoding Specificity (Tulving & Thomson, 1973) proposes that in addition to encoding the central aspects of an event, information related to the context of the event is involved. Several retrieval paths to the encoded event may exist, hence if one retrieval cue is not successful at retrieving information, another one might be successful (Tulving, 1974). The Cognitive Interview (CI), which elicits more accurate information from eyewitnesses without increasing inaccuracies (Fisher et al., 2010; Geiselman, Fisher, Cohen et al., 1986), is based on these theoretical principles. The CI is used throughout this experiment for composite construction. Of particular interest for retrieval of context is the technique (mnemonic) Mental Reinstatement of Context (MRC), which has been a part of the Cognitive Interview since its original version (Geiselman et al., 1984). Utilising the context cues to aid retrieval of memory for the to-be-remembered face is the aim in composite construction but this topic has not received much attention until more recently. Theory related to contextual cues, particularly with facial memory, and reinstating context will be outlined before describing the experiments in this chapter, which explore the impact of the mental reinstatement of context on the effectiveness (identifiability and resemblance) of composites.

3.1 Context affecting unfamiliar face recognition and identification

While earlier research on memory for words, for example, is theoretically important for the wider research of contextual cues, understanding how different cues affect memory of the main subject of interest to a criminal investigation, such as the offender's appearance, is more relevant to composite construction. Young et al. (1985) revealed how important context is for recognising faces: even highly familiar faces are not necessarily recognised if they appear in an unusual context. Also, we might confuse the identity of a person for someone else if they appear in the same context at the time of an event. This situation occurred in the well-known case for Australian psychologist Don Thomson described in Bower (1990). A woman was attacked and raped in her home in 1975. Just before the attack, she was watching a local television program in Sydney in which Thomson was (ironically) talking about eyewitness testimony. The victim was left unconscious and after she awoke, she called the police and named Thomson as the perpetrator. Thomson was arrested and on the following day the victim identified him from a line-up. He had a solid alibi though, as the TV programme at the time of the crime was broadcasted live. Thomson’s face had become associated with the event in the victim’s mind, and she confused him as the perpetrator. Due to unfamiliar face recognition being notoriously difficult to achieve accurately (e.g., Bruce et al., 1999; Burton et al.,1999), this type of face recognition is more dependent on contextual cues than familiar faces. For example, a change of background for facial images (Davies & Milne, 1982), and a change in clothing, setting and pose (Thomson, Robertson, & Vogt, 1982; Bruce & Young, 1998) all have been found significantly reduce subjects’ recognition accuracy. A similar effect has been found when hair was disguised at the time of encoding (Bruce & Young, 1998); as mentioned previously, hair, and other external features are relatively more important (cf. internal features) for unfamiliar face recognition (Ellis et al., 1979; Young et al., 1985). If identical images are used at test and study, pictorial codes (e.g., same clothing, angle of the head, lighting) are often depended on, rather than structural codes (Bruce, 1982), which are more reliable factors for the recognition of the face itself. This is one reason why, in a video-parade line-up, the suspect should not be the only one wearing distinctive items, as the aim of the line-up is to be fair and non-biased (Zarkadi et al., 2009). Conversely, for face recall, distinctive contextual cues are likely to be beneficial (i.e., to aid memory) for the face as there is no visible recognition cue to outshine other information (Smith, 1988).

3.2 From physical reinstatement of context to mentally reinstating the context

Early studies have demonstrated the effectiveness of cues with memory for words (e.g., Bahrick, 1969; Bilodeau, 1967; Bregman, 1968; Postman et al., 1955). This led to ground-breaking research on how environmental cues can facilitate memory for the to-be-remembered item. One of the pioneering studies in context reinstatement by Godden and Baddeley (1975) measured the importance of context in verbal memory. Divers learned word lists on land or under water. When retrieval occurred in the same environment as learning, rather than in the alternative environment, divers recalled significantly more words. The authors concluded that reinstating the environmental context during retrieval facilitated better recall. A meta-analysis by Smith and Vela (2001) found a reliable effect of environmental context, and that in situations where the environment was suppressed these effects were less likely to occur.

In a more recent study, Wong and Read (2011) investigated the effect of physical context reinstatement. Participants viewed a video of a simulated theft and attempted to identify the target identity in a photo line-up and recall information about the event after a 1-week interval either in the same or a different environment. The background of the video was removed so that participants would only encode the environment of the encoding environment. Wong and Read (2011) suggest that faces require relatively shallow perceptual processes compared to words, for which meaningful connections can be made, and thus face recognition may benefit from environmental cues to a greater extent. Using video of a simulated event as a study stimulus, investigating context reinstatement has been criticised to differ from a real-life event by creating two environments rather than one: that is, the environment within the video footage and the environment where the video footage itself is seen (Smith & Vela, 1992). Reinstatement of context was found to increase identification accuracy in a target present line-up. On the downside, participants' confidence level was inflated in those who were in the same room condition, which did not correspond with accuracy, while those in the different room condition with high confidence in their decision were also more likely to be correct. One potential reason for this was suggested: context reinstatement increases a sense of familiarity, which is likely to increase confidence. The accuracy of those in the different room condition could indicate that they applied some kind of mental reinstatement of context in the situation, even though they were not instructed to do so.

Earlier work by Smith (1979) found that mentally reinstating the learning environment may be almost as beneficial for retrieval as actual physical reinstatement. This concept can be easily understood by daily life examples of forgetting what you were doing on the way to pick up an item from another room for example. When you return to the room where the decision to pick something up was made, you can access this information again in short-term memory. Fisher and Geiselman (1992) describe how, if a piece information cannot be retrieved at interview, a witness can be guided to think about the environment he or she was in soon after a crime occurred, where he or she could still recall details of the original context. For this to be facilitated, the interviewer must be aware that the information has been recalled before, so as not to say anything suggestive to the witness.

A meta-analysis by Shapiro and Penrod (1986) on facial identification studies revealed a very large effect of context reinstatement on identification accuracy (d = 1.91). The difference in identification accuracy (recognition 'Hits') between participants who received context reinstatement and those did not was 27% (79% and 52% respectively). There was also a moderate effect on False alarms (*d* = 0.44). Shapiro and Penrod (1986) point out that the benefit of context is more evident in laboratory studies than in real-life studies and suggested that this could be partly due to eyewitnesses already having used many of the reinstatement cues available in a real life setting. Krafka and Penrod (1985) examined the effects of context reinstatement in a more realistic, forensically relevant setting. A target individual went to shops to purchase a small item and either 2 or 24 hours later the clerk who the target had interacted with, was asked to attempt an identification from a six-person photospread (target present and absent). Half of the participants received context reinstatement procedures that consisted of (1) instructions to recall the transaction event and mentally recreate the target's face, (2) exposure to an identification card (non-photo) and (3) exposure to another travellers' cheque containing the target's signature. In the target present photospread, participants in the context reinstatement condition identified the target significantly more accurately compared to participants who did not receive context reinstatement. In another study using more real-life relevant settings, Cutler et al. (1987b) found that mental reinstatement of context (cf. no context reinstatement), along other Cognitive mnemonics, helped to withstand the effects of disguise and biased line-up instructions in line-up identification decisions, indicating that memory of the suspect had been enhanced. In this study, additional cues for the target such as gait, voice, disguise, weapon visibility, and factors such as the degree to which the line-up members resembled the target (high similarity vs. low similarity), and line-up instructions (biased vs. unbiased), were manipulated. These additional cues increased the reliability of identifications after a two-week retention interval, as did high-similarity of the line-up members to the target.

Exposure to mugshots in the interval between encoding and identification reduced effectiveness of the line-up contextual cues, indicating that the mugshots interfered with facial memory. When a witness has been asked to construct a facial composite, factors such as how many people they have seen since the crime are out of control of a composite practitioner, but it is good to be aware of how this may affect the retrieval of the memory.

Previous research by Sanders (1984) did not find the same positive effect of physical context. Participants saw a simulated robbery on a video where a target wore distinctive clothing and glasses and were then shown a line up at five different intervals between 20 and 40 minutes. Members of the line-up either wore the same or different shirt and glasses than in the target video and were seen either in the same or a different physical setting. It was found that reinstating the context had no reliable effect on face recognition, and actually increased false positive identifications presumably due to a strong appearance cue of distinctive clothing and glasses. Participants readily chose the filler target from line-ups who was wearing the same distinctive clothing and glasses as the target. When these appearance cues were not present, filler items were selected less frequently. This indicates that distinctive glasses and clothing outshone facial information and thus were relied on more in the identification process. In their second experiment, the target was identified significantly more often than a facially dissimilar filler when the background of the line-up was a white screen. Addition of the appearance cue (from less distinctive clothes and glasses to the same distinctive appearance cues than in the stimulus video) did not increase positive identification of either the target or the filler. However, the target was picked out significantly more often as opposed to the not present response, due to the appearance cues. This underscores the fact that contextual information is important to unfamiliar face recognition; however, distinctive accessories or clothing are likely to be relied upon to a large extent, leading to misidentification. Any distinctive features, for example a tattoo on the neck, or accessories such as a branded hat, may provide useful cues for identifying an unknown perpetrator from a facial composite, even by people who have only briefly seen them. A person making an identification could for example recall seeing a hat of the same brand they were wearing, thus making a contextual connection. If a certain feature is highly unique and prominent, a composite could lead to the person being identified with a help of that feature, despite a less than desired facial resemblance to the target.

3.3 Enhancing eyewitness memory in an interview

More relevant still for composite construction, particularly for sketch composites, is how witness's memory can be enhanced during interview. In addition to the perpetrator's appearance, aspects of an event influence memory to a varying extent, and also depend on where attention has been directed (e.g., Wells & Olson, 2003). If it is possible to construct a facial composite of the perpetrator, to recover cues to their identity in the absence of other evidence than eyewitness memory, then individual factors such as the perpetrator's voice and demeanour, along with other contextual factors from the environment and psychological state of the witness, will play an important role in the Cognitive Interview (CI; Fisher et al., 1987; Fisher & Geiselman, 1992; Geiselman et al., 1984).

Guiding eyewitnesses to use cues to enhance their memory has been harnessed as one of the mnemonics in the Cognitive Interview, Mental reinstatement of context (MRC). This technique considers all contextual information and aims to enhance the main memory trace via these cues (e.g., Dando & Milne, 2009, Fisher et al., 2010). MRC is facilitated by asking a witness to explicitly think about the context of the original crime, or by asking specific questions that require a witness to think about it (Fisher & Geiselman, 1992). By drawing attention to these contextual cues is likely to enhance the primary memory, such as the face of a perpetrator. The memory record is more heavily influenced by the observer's mental thoughts than by the external environment, and therefore it is relatively more valuable to encourage the witness to focus on the former (Fisher & Geiselman, 1992). This is especially helpful when the witness has been alert and able to take in more information in an event that they are consciously trying to remember (Fisher & Geiselman, 1992). In such a situation, the interviewer should encourage the witness to recreate his/her thought processes of how he/she tried to memorise the event (Fisher & Geiselman, 1992). Factors occurring prior to the experienced event, such as life experiences, attitudes, knowledge, disposition, and biases affect how an event is interpreted (Quas et al., 2000). If new information fits well with past experiences, retention is usually better than in cases where these two factors do not fit (Quas et al., 2000). An observer might also reflect on his or her own thoughts and emotions when recalling an event, and the closer their mental state and thought processes are to the original event, the better the recollection is likely to be (Fisher & Geiselman, 1992).

Malpass and Devine (1981) examined the impact of a guided memory interview on eyewitness memory. Participants watched a staged vandalism and five months later saw photographic line-ups. They had previously seen a line-up including the target, a line-up where the target was absent, or did not see a line-up. One half of each group were given a guided interview in which their feelings, their memory for details of the room and the target, and their immediate reaction to the events were explored. After this exercise they were asked whether the target was present in the line-up. The other half of each group were simply asked this question without the prior guided memory interview. Guided memory participants were 20 % more accurate at identifying the target than those who were not guided. Participants who had seen a target in a line-up were most accurate and most confident of all the groups. The delay from target view to the line-up identification was long in Malpass and Devine (1981), and this could have increased the effectiveness of the guided memory interview, as opposed to if the identification test had followed encoding immediately. As Cutler et al. (1987b) note, a delay between encoding and retrieval stages may be required for context reinstatement to be effective. In all the experiments in this thesis the interval is approximately 24 hours, which seems a sufficient time for retrieval mnemonics to be effective.

While hypnotic-focused meditation has been found to increase accuracy of the eyewitness memory two days after encoding a staged robbery, context reinstatement outperformed this technique (Hammond et al., 2006), indicating that the environmental cues are an important part of memory retrieval. Based on previous research that has revealed that police officers do not regularly use the MRC component of the Cognitive Interview, Dando et al. (2009) designed a less complex and more succinct MRC procedure for the interviewer to facilitate. They proposed developing sketch plan (NSLEC, 2004), which can help witnesses explain what had happened, and they adjusted it to be used with MRC. This sketch plan MRC (sketch MRC) involves the witness drawing a detailed sketch or plan of the event they saw and describing this detail to the interviewer while drawing. Participants saw a video clip of a non-violent crime and were interviewed in a Cognitive Interview 48 hours later. Sketch MRC was compared to the traditional MRC (MRC) and without using MRC (no MRC). The interviews were scored for information recalled. Both Sketch MRC and MRC were found to elicit significantly more correct items than no MRC, with no marked different between the MRC conditions. Also, fewer confabulated items were found for Sketch MRC than both MRC and no MRC conditions. The sketch MRC interviews were found to be significantly shorter than MRC interviews but (as one might expect) longer than no MRC interviews. Considering the importance of this mnemonic, the important take-home from this research is that the sketch version does not reduce the effect of MRC, while reducing confabulations, and potentially makes the process more fluid. In a composite construction procedure, this technique could be ideal since it would involve a witness in the process to a greater extent (by them also sketching) and could possibly also increase motivation.

3.4 Context reinstatement in composite construction

It is normal for a composite system operator to ask the witness to think back to the crime and visualise the face (e.g., Frowd, Nelson, Skelton, Noyce, Heard & Henry, 2012), but reinstatement of context does not appear to have been explored for digital composite systems (although see Fodarella et al., 2021), nor for sketch production, and therefore this topic is of interest to this thesis. The first study investigating reinstating context mentally with composite construction, using a more realistic study design, was by Davies and Milne (1985). The participants observed the targets in an office room for just over one minute, and after one week returned to construct a composite using the Photofit system. Participants constructed a composite either in the same room where they observed the target or in a different room. They were allocated either in a guided memory procedure group or spontaneous recall group. The guided memory procedure included instructing participants to think back to the event, the environment and their mood and emotions at the time. After that, they were instructed to focus on the general appearance of the face, and then to start constructing the face. The resulting Photofit composites were evaluated by another group of participants who attempted to match them to photographs of the four targets. It was found that the composites created in the guided memory condition were matched significantly more accurately to the targets, compared to those from the spontaneous recall condition. The same environment also emerged significant compared to different environment, although with less of an effect. It was suggested that the environment had less impact in the guided memory condition, but the interaction of the factors was not significant. These findings are promising in terms of improving facial composites by a recall enhancing mental reinstatement of context mnemonic, which can be equivalent to being in the same physical environment at encoding and retrieval stages (Smith, 1979).

It has been investigated whether reinstating physical context using CCTV footage could improve facial composites using modern systems (Ness et al. 2004). A simulated crime was filmed in both high quality (used for target encoding) and poor-quality video (used as CCTV footage). In one clip, the target's back and side of the head, but not face, was visible and in another clip the target was looking directly at the camera. Each participant saw a 30 second video clip and was then assigned to one of three composite construction groups: (1) construction of a composite with no CCTV footage, (2) construction of a composite with a 7 second CCTV footage that contained the back and side of the target's head, and (3) construction of composite with the 7 second CCTV footage that contained the target looking directly at the camera. Face construction used the PRO-fit feature system. Overall, results revealed that composites constructed using CCTV footage performed better than composites constructed without this additional information. However, this effect was only present when the target's face was visible, suggesting that the incidental background information had not been encoded efficiently. It was concluded that the participants may have been fully engaged with the target (even though at the time they were not aware of the recall task ahead) rather than the environment itself.

In another experiment, Ness et al. (2006) investigated the effect of distinctiveness and context on composite construction. Increasing distinctiveness was achieved by changing the colour of the targets' shirt to bright pink. Encoding stimulus was similar to that of Ness et al. (2004), except university staff members were used as targets, who were videoed pretending to steal a mobile phone from a rucksack. Poor quality web camera footage again mimicked CCTV footage. Participants described the target's face in a Cognitive Interview and constructed a PRO-fit composite in three different groups: (1) from memory, (2) with seeing a 7 second clip of the CCTV containing the target's back and side of the head, and (3) with seeing a 7 second clip of the CCTV containing the target with his head blocked out. For two of the targets, the composites of constructors who saw CCTV footage of the target with a blocked head, were rated significantly higher than composites of constructors who saw the target's back and side of the head (no face) and who saw no CCTV footage. In addition, "no face CCTV" composites performed equally well as "blocked face CCTV" composites for one target. Composites not using CCTV only performed well for one target and not any better than "blocked head CCTV" composites, which indicates that there was a benefit of physically reinstating the context for all the targets. A new set of participants were asked to select the target from a six-person array. The results were partly in line with the likeness ratings: with the “blocked head CCTV” composites performing significantly better than the others for one of the targets. Overall, the results of this study indicate that reinstating the context by seeing CCTV footage during composite construction improves the effectiveness of facial composites.

The first study to explore the impact of mentally reinstating the context using modern composite systems was by Fodarella et al. (2021). The aim was to explore whether a more in-depth version of the mental context reinstatement (a detailed CR) procedure would improve facial composites. The procedure is described as a more "involved” use of context reinstatement by instructing participants to verbally describe the encoding environment and any thoughts and feelings they had at the time, before proceeding to recall the target’s face. The minimal CR on the other hand is a usual procedure for composite construction (in the lab and in real life) as part of the Cognitive Interview and involves instructing participants to “think back” to the event in which they encoded the target, without verbalising this information. Five experiments were conducted in total and the target pool included TV soap characters and international-level footballers. In Experiment 1, composites were constructed by EvoFIT and PRO-fit systems, and context was reinstated in three ways: Minimal, Physical and Detailed. In the physical CR, the composite construction was simply carried out in the same environment as the encoding occurred. The target encoding environment (a university cafe) was selected to be unfamiliar to participants and rich in context cues such as tables, chairs, and a small counter selling refreshments. EvoFIT composites were constructed following the procedure detailed in Fodarella et al. (2015). A standard procedure was used for PRO-fit (same as that of Ness et al., 2004, 2006). Composites constructed using a detailed CR were correctly named significantly better than both minimal CR and physical CR, and the latter two did not differ significantly. Also, EvoFIT composites performed significantly better than PRO-fit composites. These results were in line with Davies and Milne's (1985), providing support for the effectiveness of mentally reinstating the context for face construction. The difference between the composite systems in Fodarella et al. could have been enhanced, since artistic enhancement was not applied to PRO-fit composites, leaving these images looking somewhat unrealistic, while EvoFIT images are of much better quality (see composite examples in Fodarella et al., 2021, p. 185).

In Experiment 2, using a design that was similar to Experiment 1, Fodarella et al. (2021) were interested in exploring whether the Holistic Cognitive Interview (H-CI) would enhance composites further coupled with the detailed CR (cf. minimal CR). This was found this to be the case. In the H-CI, which has been consistently found to improve the computerised composites (see Frowd et al., 2015 for a meta-analysis), a witness is asked to think about the character of the target face to themselves and then make seven personality judgements (e.g., intelligence, friendliness, kindness) based on the appearance of the target face (see Frowd et al., 2008). The EvoFIT composite construction procedure had been enhanced (Fodarella et al., 2017), so that now participants were instructed to select best matches in the presented face arrays for the upper face region and after evolving the face, and to focus on all aspects of the face when enhancing the face using the software tools. EvoFIT composites were again found to be more effective than PRO-fit composites. The interaction was also significant, as detailed CR increased correct naming for EvoFIT but not for PRO-fit.

As mentioned previously, attention is key for later retrieval of memory. In Experiment 3, Fodarella et al. (2021) explored whether the contextual effect would be more impactful by directing the participants' attention to observe the environment more thoroughly: an intentional type of encoding, which involved the participants' attention being directed to the environment, by asking them specifically to pay attention to it prior to looking at the target face. The incidental encoding condition did not include such instructions. Reinstatement of context was facilitated in three ways: minimal, detailed and extensive. The first two methods were as in previous experiments, but in the extensive CR, participants were prompted or asked cued open-ended questions based on their descriptions. For example, if they had described an item in the room in general terms, they were asked if they could recall more information about it. All composites were constructed using EvoFIT from this experiment onwards due to the low correct naming rates of PRO-fit in previous experiments. No significant effect was found for CR, but there was a significant main effect of attention: intentional attention to the environment led to significantly better named composites than incidental attention. The interaction was also significant: when attention was incidental, no differences were found between the different context manipulations. Intentional attention, however, led to better-named composites in both detailed and extensive CR compared to minimal CR. Also, while intentional attention was significantly better than incidental in the detailed and extensive CR conditions, minimal CR had a similar performance in both attention conditions.

Because in Experiments 1 and 2, where environmental encoding was incidental, detailed CR improved the composites (cf. minimal CR), but in this experiment the same benefit was not replicated, and it was suggested that incidental encoding may lead to inconsistent findings due to insufficient encoding (of the environment). Fodarella et al. (2021) suggested that the engagement with the researcher might have affected the participants' attention to the environmental cues and so the interaction overshadowed environmental cues and suppressed the effect of the environment (Glenberg, 1997). Thus, in their next experiment (Experiment 4), participants were instructed to enter the encoding room before the researcher, who stayed outside until the participant had sat down. A simpler version of the mental CR was also used, according to which the participant was not guided to recall their emotional state during encoding. This revision of the mnemonic seems to go against the advice of Fisher and Geiselman (1992), who emphasise the internal state before environmental cues; however, this guidance is directed at investigative interviews in general and not specifically for facial composites. A novel way of constructing the composites was also trialled in this experiment by using a self-administered version of the EvoFIT system, which has been used for less serious crimes when police resources are limited (e.g., Fodarella et al., 2021). The design was 2 (Context reinstatement: minimal vs. detailed) x 2 (Face construction: face-to-face vs. self-administered). The encoding environment was a small office and volunteers were local residents of a small town in the UK rather than university staff and students, as was the case in their previous experiments. For the self-administered composite condition, constructors made written descriptions of the environment and the target.

Detailed CR composites were named significantly better than minimal CR. Face construction also emerged significant, with face-to-face composites named better than self-administered composites. The researchers suggest that omitting recalling the psychological state from context reinstatement mnemonic did not impair its effectiveness, which suggests that the environmental cues are more important for recalling the face. The benefit was found in both face construction procedures (and there was no significant interaction between system and CR). It would make sense for the environmental cues to support recalling faces, since both include visualisation, one of objects and one of facial details (facial features and the spacing of them). Results were similar to Ness et al. (2004, 2006), where the environmental cues from CCTV footage were found to facilitate composites.

Relevant to sketching, to examine whether the effectiveness of detailed CR is due to mechanisms based on recognition or recall, Fodarella et al. (2021) reversed the order of these two aspects in the interview. Thus, in their final experiment (Experiment 5), free recall of the face was carried out first and then context reinstatement. Two different encoding locations, both rich in environmental cues and unfamiliar to participants, were used. Minimal and detailed CR were compared, and no significant main effect was found. Researchers suggested that the fact that detailed CR did not improve composites compared to minimal CR with the reversed recognition-recall order provides evidence that detailed CR improves recall rather than recognition. This finding is encouraging for sketch composites as the procedure involves recall to a greater extent than for computerised systems. Also, this was another experiment where composites constructed face-to-face were named significantly better than composites constructed in a self-administered interview. These results were found in two experiments in Fodarella et al. (2021), which indicates that the interview process with a composite operator is an important element for the effectiveness of composites.

3.5 Context reinstatement in composite sketches

Based on these positive effects of context reinstatement on the quality of mechanical and computerised facial composites, it is worth exploring whether it could be replicated in sketch composites. Since it is indicated from the Fodarella et al. (2021) studies that the detailed context reinstatement enhances recall more than recognition in the composite construction process, it is expected that sketches will potentially benefit from this mnemonic to a greater extent. Especially standard sketching, in which no reference materials are used and is therefore even more recall oriented. The following experiments are the first to apply to sketches, and thus it will be established how mental reinstatement of context compares to physical context reinstatement in the same way as Experiment 1 in Fodarella et al. (2021). Mentally reinstating context has been found to improve PRO-fit featural composites, which is closer to sketching in technique than the holistic systems, and therefore standard sketch was compared to PRO-fit to explore whether MRC affects these systems differently.

The first experiment in the chapter was interested in exploring context reinstatement in three manipulated conditions: minimal context reinstatement, which has been part of the Cognitive Interview in the rest of the experiments in this thesis; detailed CR; and physical CR. One system, standard sketch is used. In previous research, it has been considered for the encoding environment to contain rich context cues, and therefore a more public environment was used in this experiment instead of the usual laboratory room where student participants are recruited.

3.6 Experiment 4 – The impact of reinstatement of context on sketch facial composites.

As previously, this experiment was conducted in three stages: composite construction (Stage 1), composite naming (Stage 2) and composite likeness rating (Stage 3). This time the focus of interest was how reinstating the context during a CI affects the resulting composite. Three types of context reinstatements (normal MCR used with standard sketch, physical CR and extensive MCR) were compared in a single factor design.

Methodology

3.6.1 Stage 1: Composite Construction

Design

Design in this experiment was a between-subjects single factor with three levels of types of context reinstatement: 1) minimal mental reinstatement of context (Minimal CR) 2) physical reinstatement of context (PCR) and 3) detailed mental reinstatement of context (Detailed CR). The quality of the composites was the DV, which was evaluated by a composite naming task and a likeness rating task as previously. Detailed CR has been found to lead to more identifiable composites before (see Fodarella et al., 2021) and it was thus hypothesised that it would also bring benefits for the sketch composites compared to Minimal CR and Physical CR. Participants were randomly assigned, with equal sampling, to one of these three conditions. Each participant constructed one composite.

Participants

Face constructors were first- and second-year undergraduate Psychology students from the University of Winchester who were granted course credits for their participation. They were recruited on the basis of not following the EastEnders show so they would not be familiar with the targets who were characters in this show. There were 5 males and 25 females with an age range from 18-23 years (M=19.6, SD=1.4). Equal number of participants was allocated randomly into three conditions.

Materials

Materials were 10 front-facing photographs of 5 male and 5 female actors in EastEnders and presented a neutral expression. The male targets were largely clean shaven. Images were printed in colour to dimensions of 8cm (wide) x 10cm (height). The researcher remained blind to the target photos like previously until after all interviews had been conducted and someone else prepared the stimuli according to agreed criteria.

Procedure

Participants were tested individually by the researcher (the author). They were randomly assigned to one of three conditions: 1) Minimal CR, 2) PRC and 3) detailed CR. The researcher met the participants in a university café/restaurant, briefed them about the study both verbally and by providing an information sheet to read before asking them to sign a consent form.

Target encoding

Since the stimuli were still images, the same procedure as in Experiment 3 was carried out in target encoding. This time, participants were met at a university café/restaurant, in the main area where there were more tables and people. They were asked to sit opposite the experimenter at a table for two so that nobody else was sitting right next to them. Participants were always met at the same café/restaurant to ensure that the environmental context at target encoding was the same for each condition. Participants were briefed about the study, and they signed a consent form. It was then explained that they would be looking at a face for one minute and they would need to return the next day to construct a composite of the face from memory. The experimenter explained that she needed to stay blind to the image, and was therefore going to face away from them during the target encoding. The experimenter then turned her back to the participants and they were instructed to select one image randomly from the envelope without looking through the images. The participants were asked to briefly glance at the image and, if they recognised the target, advised to put that image back into the envelope and to select another one (this occurred a couple of times). Once participants had selected a face, the experimenter asked the participants whether the face was unfamiliar, and instructed the participants to start looking at the image and started a timer. After 60 seconds, the researcher advised the participants to stop looking at the image and place it into an envelope for ‘used’ targets and to say when they had done so. Only after the image was inside this envelope, the experimenter faced the participants again. This was to ensure that the author did not get unblinded to the target by accident. Participants were reminded that the second part of the study was composite construction, and they were required to return for that the next day at an agreed time. They were thanked for taking part.

Composite construction

The experimenter and participants met 20 to 28 hours later to create a facial composite. The procedure followed the appropriate guidelines of the Cognitive Interview. Rapport was initiated between the researcher and the participants to create a relaxed atmosphere. An overview of the face-construction procedure was given after that. Those participants who were allocated to the mental context conditions were then advised to reinstate the mental context of the situation of seeing the target and the target's appearance. In the minimal context condition this was initiated by asking the participant to think back to the situation of seeing the target's image and to mentally recreate the physical environment, their psychological state and mood at the time and sensory information such as smells and sounds. They were guided also to think about what they did just before and just after the image viewing. They were given sufficient time to create this context mentally and then asked to start to describe the appearance of the target.

In the detailed CR condition, the participants were again asked to think back to the event of the target view, focus on the different aspects of contextual information such as in the minimal CR instructions, and then also to describe this information to the interviewer. They were asked to describe the context in as much detail as possible. Often participants started describing the contextual information as soon as the interviewer had finished giving them the instructions, rather than first focusing on the context in their mind. The latter also occurred with some participants. Once participants had finished describing or said that they cannot think of anything else to say, a short silence was allowed, which sometimes prompted the participant to continue his or her description. Free narrative of the target's appearance was facilitated after this. In the physical context condition the composites were constructed in the same physical environment as where the target encoding occurred. The interview procedure was otherwise the same as in the other conditions. All composites were manually sketched and followed a standard sketching process described in detail in Fodarella et al. (2015) and in Experiment 1. Sketches took about one to 1.5 hours to complete including the time for the interview and debriefing.

3.6.2 Stage 2: Composite naming

Targets used in this experiment were East Enders TV soap characters, thus people who followed the programme were recruited to the naming task.

Design

Correct naming was the DV and data was collected in a between-subjects one factorial design with three levels. The level of participants' familiarity with the targets was again assessed when they were shown to participants after the spontaneous naming round and participants were required to know at least 70% of the targets for their data to be included.

Participants

Participants who attempted to name the composites were 3 male and 33 female students from the University of Winchester or volunteers from various places in the community within 50 miles from Marlow, Buckinghamshire. They were recruited on the basis of being familiar with East Enders characters. Their age ranged from 18 to 48 (*M*=25, *SD*=10.7) years.

Materials

All 30 composites were created as grey scale sketches (pencil/charcoal/grey scale pastels) and the 10 targets were colour images. Both sets of images were printed to the same dimensions as in previous experiments. The composites were presented sequentially from one of the three conditions; each set consisted of the 10 sketches representing the targets and four foil sketches (two male, two female faces) and the composites were observed at a front view only.

Procedure

Participants were met either at the university's psychology lab, or in the community (e.g., a quiet place such as a library or their own home). It was first explained that they would be looking through some sketched faces that had been drawn by the artist from other participants' memory and descriptions, thus they were not portraits. It was confirmed that the participants were familiar with EastEnders characters and whether they follow the TV programme regularly. They were then told that their task was to attempt to name EastEnders characters (or the actors playing them) from the sketches. They were encouraged to guess if they could not think of a character spontaneously; however, they were not forced to do so. Participants were tested individually. They were given a participant information sheet to read and asked to sign a consent form before starting the task. Participants were randomly assigned to look through one of the three sets of composites created in Stage 1. Each set contained all the composites from one of the conditions. The assigned set of sketches were handed to participants (1 set per participant), who then looked through them sequentially at their own pace and offered a name (or not if they could not recognise the sketch). If they could not recall the name, other identifying factors were considered as a correct answer, if it the description was sufficiently descriptive. Participants could return to the sketches as many times as they wanted and attempt to name them again if they were unsure of one or more sketches. Once they had finished viewing the composites, the 10 target photographs were presented sequentially by the experimenter and participants were asked to name these images. The order of the composites and targets was changed randomly after each participant. The naming task took about 15 minutes to complete, including time for debriefing. Participants were thanked for their time and effort.

3.6.3 Stage 3: Composite likeness rating

Like in Experiment 3, composite likeness rating was a supplementary evaluation method for composite naming. Participants assessed the match to the target on a scale of 1-7 (1 = *poor likeness* and 7 = *good likeness*).

Design

The likeness rating scores were the DV in a within-subjects design. Participants rated all 30 composites on a scale of 1-7 (1-poor likeness, 7-good likeness) and were required to be unfamiliar with the targets.

Participants

Eighteen participants/volunteers were recruited to rate the likeness of the composites. They were students at the University of Winchester, 12 females and 6 males, aged between 18 and 54 years (*M* = 31.2, *SD* = 17.5). Participants were rewarded with course credits for participation. The volunteers did not receive a reward for participation. Criteria was that participants were unfamiliar with the target faces. Each participant rated all 30 composites.

Materials

Like in the naming task, the 30 composites were printed in greyscale and the 10 target photographs in colour, as they were seen by the composite constructors. Images were printed to the same dimensions as for face construction and naming task.

Procedure

Participants were tested individually. Each person was presented with the target picture alongside the corresponding composites from all three conditions per target and instructed to rate the similarity of each composite sketch to the target photo on a scale of 1-7 (1 = *poor likeness*, 7 = *good likeness*). They were instructed to judge how similar the composite faces looked overall compared with the target faces and not to base it on any particular feature only. It was thought it is useful for participants to see all three composites together in order to rank the likeness of them in their mind before giving the ratings for each of them. Composite-target sets were presented sequentially, in a different random order for each person, and the participants provided a rating score for each composite in a set in an unrushed manner. Testing sessions lasted for about 15-20 minutes. Participants were given a debriefing sheet to read and also verbally informed about the experimental aims.

Results

3.6.4 Composite naming

This experiment investigated the impact of mental and physical reinstatement of context on the effectiveness of sketch facial composites. There were three levels in a single factor, context reinstatement: 1) minimal mental context reinstatement (Minimal CR), 2) physical context reinstatement (Physical CR), and 3) detailed mental context reinstatement (Detailed CR). The target encoding time was 60 seconds in all conditions. It was hypothesised that Detailed CR would lead to more identifiable composites than Minimal and Physical CR based on previous findings (see Fodarella et al., 2021).

All of the targets were correctly named by all participants, and so familiarity with the target set can be seen to be ideal (*M* = 100%). Overall, spontaneous naming of composites was much higher in this experiment, at 30.6% correct (Table 3.1). Composites were correctly named similarly in Minimal CR and Physical CR, but contrary to the hypotheses, detailed CR naming was much lower than these two conditions.

Table 3.1 Percentage Correct Naming of Sketched Composites by Context Reinstatement, Experiment 4.

Context Reinstatement (s)	Sketch Composites
Minimal CR	36.7 (13.0)
Physical CR	35.0 (16.8)
Detailed CR	20.0 (10.4)
Mean	30.6 (15.3)

Note: Figures in parentheses are by-participants *SD* of the means

By-participant analysis

Univariate ANOVA revealed a significant main effect of Context Reinstatement [*F* (2,33) = 5.41, *p* = .009, η_p^2 = .25]. Simple contrasts found that Minimal CR composites did not differ significantly from Physical CR composites (*MD* = 1.7 %), (*p* = .67, *d* = 0.16), but were named significantly better than detailed CR composites (*MD* = 16.7%), (*p* = .005, *d* = 1.40). An additional t-test indicated that Physical CR composites were named significantly better than Detailed CR composites (*MD* = 15.0%), [*t* (22) = 2.63, *p* = .015, *d* = 1.07].

By-items analysis

RM ANOVA by-items revealed no significant effect of Context reinstatement [*F* (1,9) = 1.53, *p* =.24, η_p^2 = .15].

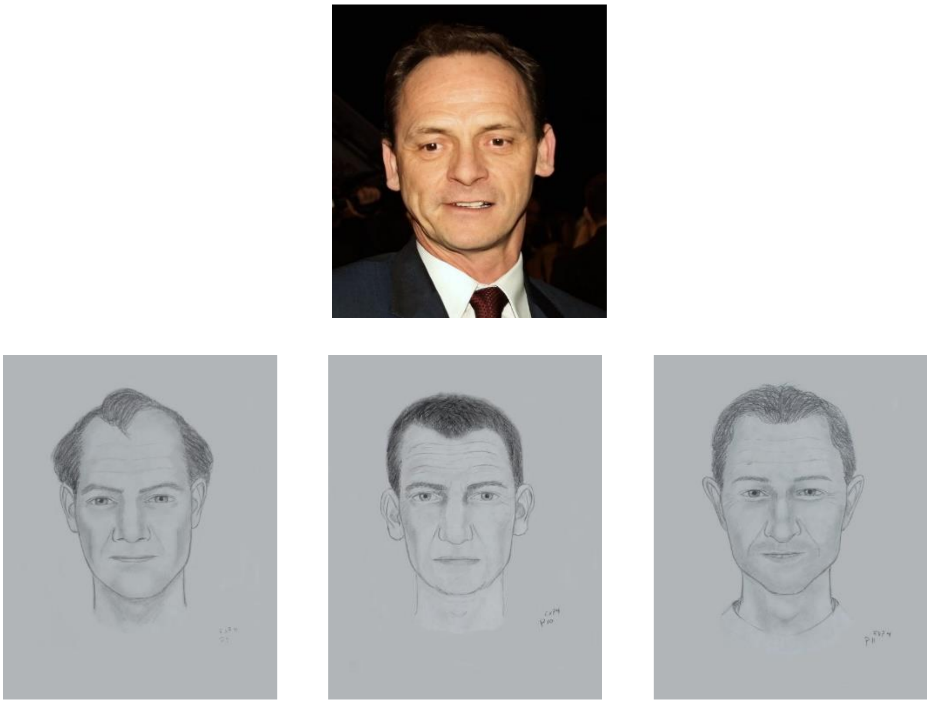
While Detailed CR did not follow expectation, such that mean naming would be much higher than the other two conditions, it appears that the naming results themselves were being driven by one or two highly performing composites, as can be seen in Table 3.2. This explains why the by-items analysis was weaker than the by-participants analysis, resulting in findings that are not in line with each other. For example, the composite of Ian Beale was named at 91.7% in Detailed CR, while most of the composites mainly had very low naming. Without this extremely well performing composite, the results for Detailed CR would have been even worse. This outcome is also illustrated by the fact that the mean statistic of Detailed CR was much higher than the median, 20 and 8.33 respectively. In the other two conditions in the experiment, composite naming was more evenly balanced. In Table 3.3, the number of times composites were named per target can be seen.

Table 3.2 Percentage Correct Naming of Sketched Composites by items.

Targets	*Minimal CR*	*Physical CR*	*Detailed CR*
Carol Branning	8.3	0.0	16.7
Kat Moon	33.3	25.0	8.3
Stacey Branning	58.3	33.3	0.0
Lauren Branning	41.7	0.0	8.3
Ronnie Mitchell	50.0	33.3	25.0
Ian Beale	41.7	33.3	91.7
Mick Carter	33.3	66.7	8.3
Max Branning	25.0	75.0	8.3
Alfie Moon	0.0	0.0	8.3
Billy Mitchell	75.0	83.3	25.0
Total Mean	**36.7**	**35.0**	**20.0**

Table 3.3 Correct Naming of Sketched Composites by items (number of times named), Experiment 4.

Targets	Min CR	Physical CR	Detailed CR
Carol Branning	1	0	2
Kat Moon	4	3	1
Stacey Branning	7	4	0
Lauren Branning	5	0	1
Ronnie Mitchell	6	4	3
Ian Beale	5	4	11
Mick Carter	4	8	1
Max Branning	3	9	1
Alfie Moon	0	0	1
Billy Mitchell	9	10	3



Minimal CR Physical CR Detailed CR

Fig 3.1 Examples of composites in all conditions with the target face from Experiment 4. The photo of the target Perry Fenwick (Billy Mitchell in EastEnders) (Wikimedia Commons, 2010), is licensed under the Creative Commons Attribution-Share Alike 2.0 Generic license. A different target image was used in the experiment.

3.6.5 Likeness ratings results

Eighteen participants rated the likeness of each composite (*N* = 30) on a scale of 1-7 (1 – *poor likeness*, 7 – *good likeness*). The resulting mean ratings (Table 3.4) were somewhat higher in the Minimal CR condition than in the Physical and Detailed CR conditions.

Table 3.4 Mean Composite Likeness Ratings by Context Reinstatement, Experiment 4.

Context Reinstatement (s)	*Sketch Composites*
Minimal CR	4.13 (0.70)
Physical CR	3.64 (0.87)
Detailed CR	3.81 (0.79)
Mean	3.90 (0.71)

Note. The Rating scale is 1 (poor likeness) .. 7 (good likeness). *SD* are shown in parentheses and are by-participant values.

By-participants analysis

RM ANOVA revealed a significant main effect of Context Reinstatement, F (2,34) = 6.53, p = .004, η_p^2 = .28. Simple contrasts of the ANOVA indicated that Minimal CR composites (M = 4.13, SD = 0.70) were rated significantly higher than both Physical CR (M = 3.64, SD = 0.87), p = .002, d = 0.62, and Detailed CR composites (M = 3.81, SD = 0.79), p = .018, d = 0.43. A t-test revealed no significant difference between Physical CR and Detailed CR ratings, p = .29, d = 0.20. Thus, Minimal CR tended to produce better quality composites overall, in particular better than Detailed CR, which is in line with the result from the naming task.

By-items analysis

The same as for composite naming by-items, RM ANOVA was not significant by-items by Context reinstatement, F (1,9) = 0.78, p = .47, η_p^2 = .80.

Discussion

Fodarella (see Fodarella et al., 2021) hypothesised that a more detailed version of the MRC would benefit the composite quality. She found that both PRO-fit and EvoFIT composites were more identifiable when the detailed CR was used compared to both physical context reinstatement and the usual, less detailed MRC. Surprisingly, an opposite effect of the detailed CR was found in this experiment study; the composites in the detailed MRC were worse than in the usual (minimal) MRC and physical reinstatement of context. Standard sketching, as explained before, relies mainly on the describer's recall and the evolving sketch. No pictures of facial features are used. This is a very different method than the computerised systems, especially the holistic systems such as EvoFIT, which focus on face recognition rather than recall. Seeing pictures of facial features/whole faces is likely to provide stronger cues than the prompting questions used in the interview to elicit more information on the face. Since context reinstatement is said to be more effective for recall than recognition (Fodarella et al., 2021), the sketching technique would have been expected to benefit from the detailed CR more.

The learning stimuli (photograph) itself could suppress the encoding of environmental cues making them less prominent in retrieval. This is called overshadowing (see, e.g., Matzel, Schachtman, & Miller, 1985), or a failure to store contextual information in memory. If contextual information is overshadowed at learning, then, according to the principle of encoding specificity, contextual cues provided at test will have no effect. Outshining, on the other hand (e.g., Smith, 1988, 1994), refers to the idea that the environment can be suppressed at test, diminishing the likelihood that ambient environmental information will be used in the construction of memory probes. Since participants were fully engaged with looking at the target image for the duration of target encoding (1 min) and not seeing a moving live target in the environment, contextual cues of the environment could have had less of an effect to memory of the primary stimulus (face) than in a real situation, where also more sensorial information might be perceived (for example the perpetrator's voice).

In addition, if the environment is familiar to some extent, it might not draw as much attention to itself at encoding, leaving the contextual cues less effective. As reported by participants during the study, the learning environment in which the participants viewed the target photo was more or less familiar to many of them but it (University restaurant/bar) had been refurbished recently, therefore it was a new environment at the same time. The effect of familiarity could have a different impact on memory potentially either improving or distorting it (e.g., Hockley et al., 2012) and using a totally unfamiliar environment for the learning stage could lead to different results. This will be considered in Experiment 5. While familiar faces do not suffer the same impairment from a changing context than unfamiliar faces (e.g. Davies & Milne, 1982), a familiar environment could provide strong contextual cues but also some different associations (from different memories of the environment), potentially impairing memory for an unfamiliar face.

Another factor affecting the learning stage is the participant's mental state. A few participants reported having felt awkward and self-conscious while looking at the photo in the busy restaurant area. No data was collected on this, only anecdotal evidence exists, therefore it is not possible to quantify the claim that this could have had an indirect impact on the composite quality. This is certainly worth looking into in the future experiments with the similar research design. Stress of experiencing or witnessing a violent crime can potentially lead to a different outcome of a composite. Hancock et al. (2011) found that when participants engaged in playing an action thriller video game while encoding the target face, the subsequent composites were recognised significantly worse than the composites of participants who merely observed the game at encoding. The heightened state of arousal in a violent crime is likely to affect the role of context also, and some contextual cues could be more prominent than others. For example, there could be a strong smell present that the witness associates with the crime at retrieval but he or she cannot remember some of the environmental context. This ought to be considered in real cases.

The overall correct naming rate was higher here than in Experiment 3 (30.6% and 19.7% respectively). This could again be partly due to the chosen target pool. This time both male and female targets were selected. It is possible that, particularly for female targets, hair was a prominent factor both in constructing the faces and recognising them, since the role of hair is important in both unfamiliar and familiar face recognition (O'Donnell & Bruce, 2001). Since sketching without reference materials relies heavily on recall, more extensive verbal context reinstatement could be subject to verbal overshadowing effect (VOE), and subsequent impairment of face recognition by verbal descriptions (e.g. Schooler & Engstler-Schooler, 1990). It was expected that detailed CR would increase verbal descriptions, since memory for the face was expected to improve (analysis of this is not included however). This was also expected to improve the composites rather than decrease their quality, and thus, the findings here are puzzling.

Some evidence exists that distinctive faces are more resistant to the VOE than more typical faces (Wickham & Swift, 2006), and therefore some impairment could be present for some targets but not others. Even if the face was more typical, but the hair was distinctive, this target is likely to be recalled better and in some cases, could lead to recognition even if the face was not constructed accurately.

Interestingly, for composites constructed with a feature system PRO-fit in Experiment 1 in Fodarella et al. (2021), the overall correct naming rate was much lower, at 9.2%, which the detailed context reinstatement increased to 13.8% (cf. minimal CR, 6.3%). The overall correct naming in the experiment, including EvoFIT composites, was 17.9%, and for detailed CR 23.8%, minimal CR 15.0% and physical CR 15.0% (Fodarella et al., 2021). The correct naming for detailed CR was therefore similar to that of this experiment, whereas EvoFIT detailed CR naming (33.8%) was close to minimal and physical CR naming here. Similar targets were used in both studies (TV soap characters), and thus it is interesting to note that the standard sketch produced composites that were named much better than PRO-fit composites, and rather surprisingly, equivalent to EvoFIT composites.

Likeness ratings were broadly in line with the correct naming results in this experiment, so that Minimal CR composites were rated significantly higher than Detailed CR composites (although Physical CR performed equivalently to Detailed CR). The ratings were also similar to Experiment 3 ratings in the 60 second encoding condition. These results indicate that the method of presenting the composites from all conditions together with the target was not unreliable. In this study, the images were presented as paper copies, while in Experiment 3 the likeness rating task was completed online with a slightly different layout: the images were presented vertically with each of the composites having the target next to them. In Experiment 3, likeness ratings did not show the same significant result as correct naming did. Online ratings may be more unreliable than in person with the experimenter, due to it being less controlled. On the other hand, it is easier to collect more data, which should reduce noise in the results.

In this experiment only standard sketch was used. Since an opposite effect of the detailed CR was found than hypothesised, and sketches were much worse in this condition compared to minimal CR and Physical CR, the next experiment compared standard sketch and PRO-fit as the construction systems. Based on the result here and in Fodarella et al. (2021), PRO-fit might perform worse than sketch in terms of overall correct naming level, but should benefit from the detailed CR. It should be considered, however, that the experimenter’s artistic skills may increase the quality of PRO-fit composites. It is rather surprising that detailed CR would have a detrimental effect on sketches, since it is a method that depends on recall the most and, as Fodarella et al. (2021) found, detailed CR enhances recall rather than recognition. Use of reference materials was not found to benefit sketch composites after a longer (60 second) encoding compared to not using reference materials in Experiments 2 and 3; however, reference materials did not hinder the process either. It is suggested that while recall is expected to improve following detailed context reinstatement, standard sketch might not be an ideal method to harness the enhanced recall for an improved likeness of the composite sketch. Reference materials could provide additional cues to achieve this. The composite system that has benefitted from the detailed CR most is EvoFIT (Fodarella et al., 2021), which indicates that a recognition part of the face construction is also improved through enhanced recall. This also provides more evidence for recall and recognition working simultaneously in face construction. In Experiment 5, adding the use of facial feature catalogues, which was found to be the better method of presenting reference materials (see Experiments 2 & 3), was tested to see if that improves the sketch composites when detailed CR is used.

3.7 Experiment 5 – The impact of reinstatement of context on sketches and PRO-fit composites.

This experiment followed up on Experiment 4 and continued exploring the impact of a detailed version of mental context reinstatement, due to conflicting findings in the previous experiment. Detailed CR has been found to lead to more identifiable composites before (see Fodarella et al., 2021), however in Experiment 4 an opposite effect was found for sketch composites. It was therefore hypothesised that the extensive mental reinstatement of context hinders composite construction by sketching decreasing the identifiability of the composites. As usual, the study involved three stages and this time the design was 2 x 2.

Methodology

3.7.1 Stage 1: Composite Construction

Design

This study was a 2 x 2 between-subjects design. The first factor was Context with two conditions: minimal mental reinstatement of context (Minimal CR) and detailed mental reinstatement of context (Detailed CR). The second factor was Composite Construction method with conditions sketch with facial feature catalogues and PRO-fit. Participants viewed the targets for 60 seconds and constructed the sketches with the experimenter 20-28 hours later. They were randomly assigned to one of the four conditions. The environment was chosen to be in a building dedicated to post graduate studies and admin, thus it was much more likely to be completely unfamiliar to the participants, which was not the case in Experiment 4.

Participants

Face constructors were first- and second-year undergraduate Psychology students from the university of Winchester. They were recruited via the university's Sona research participation system on the basis of not following the EastEnders show so they would not be familiar with the targets who were characters in this show. Course credit was given for participation. There 32 females and 8 male participants with an age range from 18 to 37 (*M* = 19.8, *SD* = 3.5) years.

Materials

The targets were ten front-facing photographs of male and female actors/actresses in Emmerdale and presented a neutral expression. The male targets were largely clean shaven. Images were printed in colour to dimensions of 8cm (wide) x 10cm (height). In the standard sketch condition pencil/charcoal/grey scale pastels and rubbers were used. And in the PRO-fit condition the initial composite was created using PRO-fit and if the participant wished to make changes to it, this was done using the various editing and drawing tools in Photoshop.

Procedure

The factors were context (Minimal CR vs. Detailed CR) and composite construction method (sketch vs. PRO-fit). The 10 targets were constructed once per condition, and thus 40 composites were created in total. Participants viewed the target face (still image) first and constructed the composite the following day (20 – 28 hr later). Participants were tested individually, and the CI was conducted as before.

Target encoding

The encoding environment was selected to be unfamiliar to the participants in this experiment. This was because, in Experiment 4, some participants were somewhat familiar with the café/restaurant where they encoded the targets. Participants were given instructions via email to come to the location, as they did not even know where in the campus the building was located. Sometimes they required additional information to find the building, such as a phone call. Participants were met at the door of a building, where postgraduate study facilities were located. The door led immediately to a postgraduate study room, where encoding was carried out. Participants were guided to sit down at the table in the middle of the room. They were briefed about the study both verbally and by being provided an information sheet to read before asking them to sign a consent form. Encoding duration was 60 seconds for all participants and the stimuli were again still images as in the previous experiments, apart from Experiment 1. Participants were told that they would be looking at a face for one minute and they would need to return the next day to construct a composite of the face from their memory, which would be carried out on the main campus. The experimenter explained that she needed to stay blinded to the image and was therefore going to face away from them during the target encoding. The experimenter then turned her back to the participants and they were instructed to select one image randomly from the envelope without looking through the images. The participants were asked to briefly glance at the image and, if they recognised the target, advised to put that image back into the envelope and to select another one. This occurred on 2 occasions. Once the participants had selected a face that was unfamiliar, the experimenter confirmed this was the case by asking the participants and instructed the participants to start looking at the image and started a timer. After

60 seconds, the researcher advised the participants to stop looking at the image and place it into an envelope for 'used' targets and to say when they had done so. Only after the image was inside this envelope, did the experimenter face the participants again. This was to ensure that the author did not get unblinded to the target by accident. Participants were reminded that the second part of the study was composite construction, and they were required to return for that the next day at an agreed time. They were thanked for taking part.

Composite construction

The researcher and participants met 20 to 28 hours later in one of the research laboratories (different location to encoding environment) to create a facial composite. Cognitive Interview was conducted as before. Participants were instructed to reinstate the context mentally in one of the two ways (Minimal CR or Detailed CR), as described in Experiment 4, before the researcher proceeded to facilitate free recall. The sketch procedure was carried out as in previous experiments with an option to use facial feature catalogues and conditions using PRO-fit proceeded similarly to Experiment 1. If the participant wanted to make more extensive edits to the PRO-fit image, this was carried out in Photoshop using a digital drawing tablet. With sketches, the procedure using facial feature catalogues was similar to Experiment 3 and occurred after the initial sketch had been created and developed further by following participants' instructions. Unlike in Experiment 3, sketching was carried out manually on the paper, as opposed to digitally, and therefore no pictures of the facial features were taken but the features selected by the participants were copied by hand from the catalogue. Sketches took about 76 minutes to complete and PRO-fit composites took about 51 minutes.

3.7.2 Stage 2: Composite naming

The main evaluation method in this study was composite naming, which was completed with participants who were appropriately familiar with the target pool (Emmerdale TV soap characters).

Design

The DV was correct naming and collected in a 2 (Context) x 2 (Composite construction method) between-subjects design. The level of familiarity with the targets was assessed as before, by showing the target photographs after naming of the composites; as normal, participants were required to name at least 7 of the 10 targets.

Participants

Volunteers were firstly recruited from people known to the researcher. The criteria for volunteers were that they follow or have recently followed the Emmerdale soap programme regularly enough to know its characters well. More people were recruited from Emmerdale fan groups from Facebook. These were people unknown to the researcher. In addition, a research assistant from the University of Central Lancashire recruited more volunteers from people she knows. There were 25 male and 7 female volunteers, and their age ranged from 18 to 72 (M = 44.1, SD = 16.3) years.

Materials

Materials were the 20 greyscale composite sketches (10 from each method of construction), 20 PRO-fit composite images and the 10 target colour photographs. Each composite set included four foil images as before. Images were seen by the volunteers either in a face-to-face meeting or remotely on a Skype call (audio). The images in the face-to-face viewing were printed to the same dimensions as in previous experiments. In the online viewing, the images were shown in their original digital file size: 350 dpi, width: 2952 pixels, height: 4121 pixels for sketches, and 96 dpi, width: 384 pixels height: 576 pixels for PRO-fit composites, via the researcher's shared computer screen.

Procedure

Two methods of the naming task were used. Because suitable participants were based around the UK, the author conducted the naming task remotely via Skype audio call. A research assistant did this face to face with volunteers in the same physical space and proceeded as was the case in previous experiments. In the naming task via Skype, the researcher agreed a meeting time with the volunteers via email or Facebook messenger. They then logged onto Skype (or first installed Skype if they did not have it yet) and a call was made by the researcher using audio only. The researcher then explained the procedure briefly, showed the briefing sheet via screen share option on her computer screen and sent the consent form to the volunteers via email who signed it and sent it back to the researcher. It was explained to the volunteers that the composites were constructed by the researcher from other participants' memory and descriptions of Emmerdale TV soap characters either by sketching or with the PRO-fit composite system. And thus, not to expect the images to be portraits of the characters but rather best likenesses that the participants could achieve. The researcher then started showing the composites from a randomly selected condition. These were either sketches or PRO-fit composites. The composites of each condition were also shown in a random order each time. The composite images were selected separately from a folder, where all composites for one condition were saved into after they had been constructed. The folder was open in the background, but covered by the composite image every time a new one was clicked on to show them as a bigger image. The volunteers set the pace in which they saw the composites by indicating when they wanted to see the next one or if they wanted to see all or certain composites again. The researcher communicated with participants throughout this stage, as sometimes they tended to not indicate clearly when they were ready for the next image. Participants were advised that they may see any of the images again and to say so when this was the case so that the researcher could note down the composite number. When the volunteers had finished attempting to name the composites, the target faces were shown to them in a randomised sequence, and the volunteers were asked to name them. The same a priory rule applied as before: volunteers were required to be familiar with at least 7 targets for their naming data to be included.

3.7.3 Stage 3: Composite likeness rating

Naming data provided useful data for this experiment. As before, evaluation by likeness ratings was also included. Participants assessed the match of sketches to the relevant targets on a scale of 1-7 (1 = *poor likeness* and 7 = *good likeness*).

Design

The likeness rating scores were the DV in a within-subjects design. Participants rated all 40 composites on a scale of 1-7 (1-poor likeness, 7-good likeness) and were required to be unfamiliar with the targets. Rather than meeting face to face, participants completed the task on Qualtrics survey website.

Participants

Thirty participants were recruited to rate the likeness of the composites. They were students and staff at the University of Winchester, 23 females and 7 males, aged between 18 and 62 years (M = 25.7, SD = 10.6). Participants were rewarded with course credits for participation. Criteria was that participants were unfamiliar with the target faces. Each participant rated all 40 composites.

Materials

Qualtrics online survey platform was used to design the likeness rating survey. Participants were instructed to use a bigger tablet or a laptop/desktop to complete the task on and not use a mobile phone due to the small screen size. A link to the survey was provided via the university's research participation site Sona.

Procedure

As in Experiment 3, the likeness rating task was conducted on Qualtrics survey website. Each facial composite from each of the four conditions were presented with their matching target image and laid out to be viewed one after another. The participants were required to give a rating to each composite image from 1 (*poor likeness*) to 7 (*good likeness*) and then proceed to the next target and its composites.

A box for the rating number was located under each set of images (composite and target image). Only a number code of the composite was used for the identification of each composite. Participants were also asked to write the name of the player after each set of four composites if they recognised them from the target image (this did not occur in any cases). The survey was designed so that participants could not proceed before all questions had been answered (to avoid missing ratings).

Results

3.7.4 Composite naming

This experiment continued to investigate the impact of context on composite effectiveness using two different composite systems. In Experiment 4, context was reinstated physically and with two different mental CR methods, minimal and detailed. Contrary to expectations, the detailed CR had a detrimental effect on the composites and thus a replication study is required. This is especially since the encoding environment was familiar to some extent to some participants in Experiment 4, which may have affected the results. This time, only the mental CR conditions were tested, and composites were constructed either as a standard sketch or a PRO-fit composite.

The targets were named very well, a mean of 95% correct. Due to some un-named targets, conditional naming was used in the analysis (see Experiment 1, Results). Overall correct spontaneous naming of the composites was the highest so far in all experiments, at 51.5% (Table 3.5). The naming levels were similar in all conditions apart from in the Detailed CR sketch, which was much higher. The number of times composites were named for each target is shown in Table 3.6. Examples of all composites and their target image shown in Figure 3.2.

Table 3.5 Percentage Correct Conditional Naming of Composites by Context reinstatement and Composite system, Experiment 5.

Context reinstatement	*Composite System*		
	PRO-fit	Sketch	Mean
Min CR	49.5 (21.6)	47.3 (16.9)	48.4 (18.8)
Detailed CR	43.9 (24.0)	65.3 (11.0)	54.6 (21.1)
Mean	46.7 (22.2)	56.3 (16.6)	51.5 (19.9)

Note: Figures in parentheses are by-participants *SD* of the means.

Table 3.6 Correct Naming of Composites by items (number of times named), Experiment 5.

Targets	Min CR sketch	Min CR PRO-fit	Detailed CR sketch	Detailed CR PRO-fit
Debbie Dingle	5	5	2	3
Charity Sharma	3	5	5	1
Chas Dingle	5	5	5	5
Bernice Blackstock	6	4	2	2
Moira Dingle	5	6	7	6
Robert Sugden	0	5	2	4
Aaron Livesy	5	6	6	2
Ashley Thomas	3	4	2	2
Cain Dingle	4	7	6	7
David Metcalfe	1	3	1	1

By-participant analysis

IS ANOVA revealed no significant effect of Context reinstatement [F (1,28) = 0.84, p = .37, η_p^2 = .03]. No significant difference was found for Composite system [F (1,28) = 2.03, p = .17, η_p^2 = .07], but the interaction approached significance, [F (1,28) = 3.08, p = .09, η_p^2 = .10]. A simple main effects analysis revealed that the benefit of context reinstatement (Detailed CR > Minimal CR) was only present for sketch and not for PRO-fit (*MD* = 21.4%, p = .033); while PRO-fit performed slightly better in the Minimal context reinstatement condition (*MD* = 2.2%, p = .82).

By-items analysis

RM ANOVA revealed no significant effect of Context reinstatement [F (1,9) = 1.52, p = .25, η_p^2 = .15] nor Composite system [F (1,9) = 2.75, p = .13, η_p^2 = .23]. Similar to the by-participant analysis, the interaction approached significance [F (1,9) = 4.16, p = .07, η_p^2 = .32]. A simple main effects analysis revealed that the benefit of context reinstatement (Detailed CR > Minimal CR) was only present with sketch and not with PRO-fit (*MD* = 15.6%, p = .014); Minimal CR did not have this effect for PRO-fit (*MD* = 1.6%, p = .98).

Min CR sketch   Detailed CR sketch   Min CR PRO-fit   Detailed CR PRO-fit

Fig 3.2 Examples of composites in all conditions with the target from Experiment 5. The photo of the target Charley Webb (Debbie Dingle in Emmerdale) (Wikimedia Common, 2015), is licensed under the Creative Commons Attribution 2.0 Generic license. A different target photo was used in the experiment.

3.7.5 Likeness ratings

Thirty participants rated the likeness of each composite (*N* = 40) on a scale of 1-7 (1 – *poor likeness*, 7 – *good likeness*). The resulting mean ratings (Table x) were higher for the sketch composites, but also (curiously) when Minimal CR was used (cf. Detailed CR), which is the opposite finding than in composite naming.

Table 3.7 Mean Composite Likeness Ratings by System and Encoding Duration, Experiment 5.

Context Reinstatement	*Composite System*		
	PRO-fit	Sketch	Mean
Minimal CR	3.03 (0.87)	3.83 (0.85)	3.43 (0.79)
Detailed CR	2.84 (0.79)	3.44 (0.82)	3.14 (0.75)
Mean	2.94 (0.79)	3.64 (0.78)	3.29 (0.73)

Note. The Rating scale is 1 (poor likeness) .. 7 (good likeness). By-participants SD values are shown in parentheses.

By-participants analysis

RM ANOVA revealed a significant effect of Context Reinstatement, [F (1,29) = 10.14, p = .003, η_p^2 = .26]. Simple contrasts indicated that Minimal CR composites were rated significantly higher than the Detailed CR composites (MD = 0.29). The effect of Composite system was also significant, [F (1,29) = 43.56, p < .001, η_p^2 = .60], with sketches emerging with significantly higher ratings than PRO-fit composites (MD = 0.70). The interaction was also found to be significant, [F (1,29) = 4.56, p = .04, η_p^2 = .14], and simple contrasts revealed that Minimal CR was superior to Detailed CR for sketch (MD = .39, p <.001, but only approached significance for PRO-fit (MD = .19, p = .06).

By-items analysis

RM ANOVA revealed no significant difference in the ratings of the different context conditions [*F* (1,9) = 2.23, *p* = .17, η_p^2 = .20]. The effect of Composite system was significant, [*F* (1,9) = 10.17, *p* = .011, η_p^2 = .53], with sketches being rated higher than PRO-fit composites (MD = .70). The interaction was not significant, [*F* (1,9) = 0.23, *p* = .64, η_p^2 = .03).

Discussion

In Experiment 4, the detailed CR was found to decrease the effectiveness of the composites, which was not expected based on previous studies investigating the impact of context in facial composite construction (see Fodarella et al. 2021), which have found that both PRO-fit and EvoFIT composites are improved by the detailed CR. In this experiment, however, the results indicate that the detailed CR can enhance sketch composites. Although there was no significant main effect of context in composite naming, interaction approached significance because detailed CR improved sketch composites compared to minimal CR but this was not the case for PRO-fit. In Experiment 2 in Fodarella et al. (2021), the same occurred: Detailed CR did not improve PRO-fit composites unlike it did for EvoFIT composites. It is possible that the sketching method, in which the initial sketch is based on recall, benefits from the detailed CR particularly at this first stage of the sketch construction. Reference materials (selected from facial feature catalogue) are then likely to enhance the likeness of a composite further when a participant is able to refine the facial features by finding features that look similar to their mental image of the target's features.

The target pool in this experiment was similar to that of Experiment 4 in size, and both contained female and male targets, and thus, the correct composite naming results can be more directly compared. Overall correct naming was again higher than in Experiment 3 (51.5% and 19.7% respectively) and also higher than in Experiment 4 (30.6%). It is evident in the number of times each composite was named by participants (see Table 3.6) that most conditions had many highly identifiable composites and that only one condition (minimal CR sketch) had one composite (of Robert Sugden) which was not identified by any of the eight participants. In comparison, in Experiment 4, all conditions had composites that were not identified by anyone, and one condition (Physical CR) had three of these. The level of distinctiveness was not measured, but it is possible that some distinctive features would have aided recognition for certain targets. For example, there was a prominent fringe and dark hair for the target Chas Dingle. The composites of this target were named by five participants (out of eight) in each condition. In comparison, two of the likeness ratings for these composites by items were below average and two above average (of total rating average for all composites). This indicates that hair is likely to have played a part in making the composites more recognisable.

Based on the results of Experiments 1 and 2 in Fodarella et al. (2021), PRO-fit composites were expected to be named much worse than they were. In fact, in this experiment, the average correct naming of PRO-fit composites was 46.7%, and no difference for sketch and PRO-fit was found overall. The researcher used Photoshop to edit the initial PRO-fit images, firstly to fill in any gaps in the images if a feature on PRO-fit did not fit the rest of the face perfectly, so that the face looked as realistic as possible. Often, quite extensive editing was also carried out for many composites. For example, hair was drawn using the pen and paintbrush tools to add more detail to it. The researcher applying her artistic skills therefore is likely to have affected the outcome, as has been found with enhanced Photofit composites (Gibling & Bennett, 1994).

Naming of composites was carried out by two different methods in this experiment: remotely via a Skype audio call and face to face. No pilot study of the remote method was conducted; therefore, it is possible that this could have led to some differences in how participants named the composites. In the face-to-face method, participants were physically holding the A4 papers that presented one composite each and had full control of the pace at which they looked through the images. They were able to also physically set one or more composites aside and return to attempt to name them after viewing all composites. On a Skype call, the researcher shared her computer screen with participants to show the composites to them. She instructed the participants to say when they were ready to move on from one composite to the next, but often had to ask the participants if they had not indicated this clearly. It is possible that this could have applied pressure for the participants to make a decision, even though the researcher was careful not to rush participants. The researcher thought that it was best for the composites to be presented in this way as she had to select the images from a folder individually in a random order. If it was logistically possible to randomise the composite order for each participant, it may be better to send the composite folder to them, which would allow participants to view the composites themselves. This would allow them to have full control of the pace, as is the case for face-to-face naming.

There was an aspect of familiarity encoding the environment in Experiment 4, which could have played a role in the suppression of environmental cues for those participants, who were more or less familiar with the university restaurant (although it had been refurbished recently). It was made sure that the encoding environment was totally unfamiliar to participants in this study for that reason. The building was also away from the main campus and the participants often had to make more effort to find the place, which could have made the contextual cues more meaningful. Curiously, the likeness ratings indicated that the minimal CR was overall better than the detailed CR such as was the case in Experiment 4. And a significant interaction revealed that the minimal CR improved sketches but not PRO-fit composites. Sketches were overall rated better than PRO-fit composites. The likeness rating data was collected on an online survey and since the researcher does not oversee the process, there is a question whether ratings are affected by the data collection method and if so, which way. This will be discussed more in the general discussion chapter.

Summary of the experiments

It was hypothesised that since the process of constructing sketch composites is very much recall oriented, the detailed mental reinstatement of context (MRC) should also improve sketch composites. Two experiments were conducted. Experiment 4 manipulated one factor, context, at three levels including the usual MRC (witness thinks about the context of the event but does not verbalise this), detailed MRC and physical context reinstatement (composites constructed in the same environment where encoding occurred). Opposite to expectations, the composites following the detailed MRC performed significantly worse than the other context conditions. There was an element of familiarity in the encoding environment to some participants, which might have affected the results. In Experiment 5, a completely unfamiliar environment was chosen for target viewing. The design was 2 x 2; context (usual and detailed MRC) x composite system (PRO-fit and standard sketch). More in line with expectations, the detailed MRC improved the standard sketch composites, but this effect did not extend to PRO-fit composites. Participants encode the environment in a unique manner, which may lead to inconsistencies in the effectiveness of the context cues. Overall, the results indicate that in the standard sketching process, which is more recall oriented than constructing a PRO-fit composite, more focus on the contextual cues at retrieval led to enhanced recall and thus improved composites.

4 CHAPTER 4 – DISCUSSION

Despite sketching being the oldest method of composite construction (e.g., Taylor, 2001), the technique itself has not been a focus of research until more recently (e.g., Kuivaniemi-Smith et al., 2014). It is usually likened to feature systems such as E-FIT and PRO-fit (e.g., Frowd et al., 2005b); however, the procedure does appear to include a more holistic element when the initial sketch is not revealed to the witness until the whole face has been lightly drawn (e.g., Davies & Little, 1990; Fodarella et al., 2015; Taylor, 2001). And sketches have been assessed to perform more effectively than other feature systems when the delay from seeing a target face to composite construction is a day or more in duration (Frowd et al., 2005a; Frowd et al., 2015). Sketch artists are still utilised across the world, especially in the USA, and there are circumstances in which only a sketch artist can currently produce a composite, for example when the witness has only seen the offender in a profile view or another angle differing from a front view (as computerised composite systems generally only construct the face in this view). This thesis has started to address gaps in the literature on the sketching technique, with key questions such as whether reference materials were beneficial in the composite construction process and for how long a target face had been encoded. The aim was to develop the technique to produce more effective (more identifiable and bearing a closer resemblance to the target face) sketch composites.

In Chapter 2 experiments, the main focus was to investigate the impact of reference materials on the resulting composites when the target encoding duration was relatively short (5 sec) and longer (30 sec and 60 sec). One of the main findings was that a short encoding duration leads to less effective composites (less frequently named and that resembled the target less), indicating that the memory of the face is weaker than after longer encoding. This was expected based on similar findings, most of which is centred on face recognition (e.g., Light et al., 1979; Memon et al., 2003; Shepherd et al., 1991). The interesting question connected to encoding duration was whether using reference materials in sketch composite construction can be beneficial in producing better composites. Most sketch artists use reference materials (e.g., Kuivaniemi-Smith et al., 2014; Taylor, 2001), but there are some artists who do not, in particular due to the concern that they may interfere with a witness's memory of facial detail (e.g., Boylan, 2000; Zamora, personal communication, 2017).

4.1 What worked and what did not with reference materials?

4.1.1 Benefit from the reference materials after short encoding

No overall benefit of reference materials was found in the three experiments in Chapter 2. However, for shorter encoding this was apparent, both using PRO-fit as the main system and using a facial feature catalogue in the sketching procedure. In Experiment 1, 5 second PRO-fit composites were much better than 5 second sketches, and the opposite result occurred in the 30 second conditions. Also, while PRO-fit performed similarly regardless of encoding time, the short 5 second encoding impaired sketches compared to 30 second encoding. This indicates that the standard sketching technique, which relies heavily on recall and in which the only visual point of comparison is the evolving sketch, is not able to sufficiently support the retrieval of memory of the face on its own. Participants sometimes stated that their mental image of the target had started to fade or was almost completely gone after they first saw the initial sketch, after free recall and probing questions on facial features. There could be an element of verbal overshadowing (e.g., Engstler-Schooler, 1990) involved in this technique. This phenomenon is said to occur when verbal codes that are generated when an unfamiliar face is first being encoded interfere with later recognition (Wickham & Swift, 2006).

In Experiment 3, a facial feature catalogue was used after the sketch had been developed into a face containing all facial features, and only then shown to participants. The overall correct naming was still far from ideal, with a mean of 19.7%, the best naming (26.8%) emerging with 60 second sketches with no reference materials, which can be said to be reasonably good for composites in general (see Frowd et al., 2015). Selecting facial features from a catalogue was beneficial in the 5 second condition (22.1% correct naming), which performed significantly better than standard 5 second sketch, but in the 60 second condition they did not improve the composites further compared to standard sketch. In fact, the catalogues enabled achieving composites that were equivalently named to the 60 second composites in both construction systems. This indicates that the memory of the face was intact even after such a short duration of encoding and that reference materials helped to retrieve the finer facial detail after a whole face had been roughly developed as a sketch. This suggests that the initial stage of a composite sketch, during which the witness is asked more probing questions about the facial features, is important for retrieval rather than detrimental. When a witness first sees the sketch as a whole face, they either recognise or do not recognise it as the face in memory. Or certain features may look familiar, however, but the wrong configural information may be present, which makes recognition more difficult. Since the initial sketching stage was also focused on the spacing of the features without the witness seeing the face yet, it could also be the case that this information is recalled accurately, but the features themselves are not accurate. It would likely be in this situation that the composite sketch benefits from the witness viewing some reference materials to achieve better likeness in the features. By the witness viewing the holistic sketch and pictures of facial features from a catalogue together combines featural and holistic processing, which is likely to lead to a more accurate composite. While there is certainly scope for optimising the sketch composite process to further improve naming, Experiment 3 findings indicate that the procedure of first drawing the face from a witness's recall, prompting more information by mainly open-ended questions, and then presenting the sketch as a whole face to this witness, is compatible with the later focus being on pictures of individual facial features.

This interconnected nature of holistic and featural processing of a face has previously been demonstrated. Cabeza and Kato (1998) challenged the theory that emphasises configural processing in an extreme or moderate way by generating prototype faces that contain either configural or featural information from four different faces. The configural prototype consisted of the four faces morphed together both in shape and in colour-texture information, and featural prototype used features from all the different faces in one face: the outline and cheeks of the first face, the eyes and eyebrows of the second face, the nose of the third face, and the mouth of the fourth face. In the configural prototype, configuration was maintained but features distorted, whereas in the featural prototype, this was the other way round. The experiment included a study and test phase. Each face was studied for 3 seconds. The test included either studied prototypes, absent prototype (i.e., the exemplars were studied but the prototypes were not), new prototypes (i.e., neither exemplars nor prototypes were studied; the new prototypes came from two different sets), studied exemplars, and new exemplars (from two different sets). Each face was presented for 4 seconds, and participants were required to rate it based on how familiar they were with each face (1 - sure new, 2 - probably new, 3 - probably studied, 4 - sure studied). The results offer evidence that both configural and featural information is important. By falsely recognising faces, that consisted of features seen in different (studied) faces but presented in a new configuration, this indicates that the featural information had been stored in the participants' memory. Also, the prototype effect was as strong for prototypes with completely new configurations as for prototypes that maintained the global configuration of studied faces, which further demonstrates that featural processing has a bigger role than the theory that emphasises configural processing of a face perhaps gives credit for.

Different composite systems also potentially have more common features than often thought. EvoFIT is a holistic system with which internal features and specifically the eye region is constructed first (e.g., Fodarella et al., 2017; Frowd et al., 2012; Skelton et al., 2019). Its performance has been found to be high after 10 second encoding of a target's face (esp. without a weapon being present at encoding). It appears, that the process of EvoFIT, PRO-fit and sketch that use reference materials after a whole face has been drawn from recall all share elements of the holistic processing of a face (e.g., Tanaka & Farah, 1993; Tanaka & Sengco, 1997), which is likely to be more prominent after short encoding duration (e.g., Goffaux & Rossion, 2006). With sketch, this needs to be achieved through more extensive recall first. In other systems too, face construction starts by asking a witness to describe a face; however, face recognition plays a bigger part since the photographic face is visible to the witness after this point. With PRO-fit, some of the descriptions are entered into the system to narrow down the pictures of each facial feature to a manageable number, before showing the whole face to the witness, who then starts to select features that resemble the target better. On EvoFIT, more general descriptions are entered, and witness then proceeds to select faces from already whole faces earlier in the process, which is likely to be a more effective method when memory is weak, compared to having to rely on recall, which is usually more difficult to achieve than recognition (e.g., Bruce et al., 1999; Burton et al., 1999).

4.1.2 Why did selecting facial features in PRO-fit not benefit the sketching process after short encoding?

In Experiments 1 and 2, composite construction started similarly in the reference materials condition. After free recall, the researcher narrowed down the pictures of facial features based on the participant’s descriptions, as described in Fodarella et al. (2015). While participants continued working on PRO-fit in Experiment 1, in Experiment 2 PRO-fit was used for participants to select facial features (in a whole face context) after free recall and the composite was then continued as a sketch, and thus all composites were produced as sketches at the end of the interview. This time, no interaction between encoding duration (5 sec. vs. 60 sec.) and construction method (PRO-fit ref materials vs. standard sketch) was observed, and using PRO-fit to provide reference materials for the participant did not improve the composites compared to standard sketching. This outcome was unexpected, as face processing is likely to be more configural in nature after short encoding, due to the lack of time for each feature to be encoded in a detailed manner (e.g., Goffaux & Rossion, 2006). Selecting PRO-fit features in a whole face context should have, in theory, supported holistic processing of a face, and thus, led to better composites.

After participants had selected facial features in a whole face context in PRO-fit, composite construction was continued by sketching, which required initially copying the PRO-fit image on paper by drawing. While the researcher is sufficiently skilled at drawing portraits with good likeness (which this drawing stage required), the composite interview is very different to carefully portraying a face and requires fast-paced drawing to maintain witness's motivation in the task. Naturally this phase took more time than normal (est. to be about 10 minutes longer) and might have contributed to lowering motivation, inhibiting the effectiveness of the final image. The researcher's visual judgement on how similar the PRO-fit image and the sketched face were, was relied on rather than a more measurement-based method. This could have introduced some errors in the initial composite sketch. This preposition was considered in Experiment 3, in which a digital drawing method was used. Pictures of facial features, that were selected by a participant, were overlayed on the sketched face on a new layer in the Procreate drawing programme, for more accurately copying features into the composite face. It seems reasonable to suggest that any kind of method that reduces subjectivity may improve the outcome of a composite.

The mismatch of the photographic face and a sketched face that contained relatively little shading, particularly in the early stage of sketching, is likely to have affected participants' visual perception and possibly interfered with their memory of the face. Faces are not recognised well by the outline of facial features alone, and surface characteristics such as pigmentation and texture of skin and hair provide more useful information (Bruce, Green, & Georgeson, 2003). Photographic images contain this information to a greater extent than sketching. Line drawings are missing this detailed information altogether, and even faces with a level of familiarity (celebrities) are poorly recognised as such drawings (Davies, 1978). When drawings contain shading information, recognition is better (e.g., Bruce et al., 1992; Davies et al., 1978; Rhodes et al., 1987) by providing more shape cues to the face (Bruce, 1988, 1989; Davies, 1978). It is worth noting that the shading information may be represented differently in targets of darker skin tone and that these targets should naturally be included in future research. Sketches in this thesis were drawn with sufficient shading information to be realistic looking, but nevertheless contained less texture information than composites from computerised systems (Frowd, 2012). While this could be beneficial to recognition by reducing incorrect information (Frowd, 2012; Frowd et al., 2008), for witness's memory of the face this could cause interference by first seeing detailed facial detail in PRO-fit and then less detail in a sketch, and some surface information that the witness might have found useful in the PRO-fit composite, may have been lost when the face was sketched by copying the initial PRO-fit composite.

It would also be sensible to compare the best sketching procedure in constructing a composite after short encoding to a holistic composite system. Since EvoFIT has been developed by extensive research for years, and this thesis is in the author's knowledge the first one to address the sketching technique itself in a detailed manner, it would have perhaps been too early to compare these two systems without investigating sketching first. It is possible to construct EvoFIT composites, that are well named after a 10 second encoding of the target face (Erickson et al., 2021), which again points to the importance of configural processing of the face after short encoding. The aim is to find out what works best in facilitating witness memory in different circumstances and ultimately recommend the best composite system to use in each case.

4.1.3 How could the reference materials be optimised?

Some participants reported that their mental image started getting less clear after seeing many features in the catalogue. This could indicate an element of interference to memory, and it would be worth exploring different ways of presenting reference materials to a witness. Taylor (2001) emphasises that a witness should be guided to use catalogues rather than looking through any sections containing hundreds of pictures. This recommendation was followed in Experiments 3 and 5; however, perhaps a stricter protocol for this could be followed. For example, witnesses could be instructed to only focus on the eye-eyebrow-region, since the upper facial features appear to be more important for face recognition (e.g., O'Donnell & Bruce, 2001). The current EvoFIT procedure also instructs witnesses to focus on the upper half of the face, since this has been found to lead to more identifiable EvoFIT composites (Fodarella et al., 2017).

The initial sketch being presented as a whole face appears to be an important step in achieving reasonably effective composite sketches. This has also been found by Widden et al. (2017), who compared a manual sketch and a photographic sketch, where facial features from a facial feature catalogue were edited into the sketch in Photoshop. Since in Experiment 3, using a facial catalogue was found to be helpful in the 5 second encoding condition after the sketch had been established as a whole face, it could be tested if whole face reference pictures could be more compatible with viewing the developing holistic sketch. In some facial feature catalogues, such as the FBI facial identification catalog, the whole face is visible, and using this could be compared with the Steinberg's catalogue that presents individual features while the rest of the face is blocked out with a shape. As discussed in the literature review of Chapter 2, some sketch artists utilise pictures of whole faces, and while this risks interference to memory, introducing these pictures after the main features and look of the face have been established, is likely to prompt refining of the features rather than making drastic changes. But again, this is likely to be dependent on the strength of memory and a weaker memory is likely to be more vulnerable to a witness becoming overwhelmed by different visual cues in other faces.

PRO-fit was used to select facial features in a whole face context after free recall in Experiment 2. It would not be possible to switch the order and use PRO-fit after the initial sketch has been created since it would be difficult to find exact matches for features that had already been established as a light sketch.

It could also be investigated whether basic shapes of facial features, rather than actual images of them could offer some cues for remembering the shapes of the target's features. A type of preliminary catalogue to support the vocabulary for describing the shape of the eyes or face, for example. Using such materials could also prevent misunderstanding between the witness and the artist during initial recall because describing a face as square, for example, may mean rectangular to someone else. This method could also potentially eliminate interference from other faces' photographic features. It is predicted, though, that more detailed reference materials might be needed at a later stage of sketch composite construction if this type of catalogue was used at the beginning.

According to Frowd (in Wilkinson & Rynn, 2012, p.42-56), sketches containing less detail than composites produced with computerised systems, appears to be due to difficulty in witnesses remembering the texture of the face. This leads to areas of the composite being left blank or shaded only lightly, which may prove problematic for subsequent identification. For this reason, future research is suggested to explore combining two different methods that do not differ as much in visual detail as sketch and PRO-fit does. There is, in fact, a pilot sketch database for EvoFIT (Lampinen, Erickson, Frowd, & Mahoney, 2015), that contains hand drawn images rather than photographic ones (see Figure 4.1) and this could be trialled with sketching in a similar way to Experiment 1. By constructing internal features effectively of a “computerised sketch”, this could retain some of this information by generating the sketch-like image quicker, which could then be built upon if requested by the witness.

4.2 Effects of the target pool on composite naming

4.2.1 Potential reasons for low correct naming rate

In Experiments 1 and 2, the correct naming rates were so low that inferential analyses were not included. As discussed in each experiment, one cause for low naming could be to do with the target pool: how many potential identities the evaluator is required to search for from their bank of familiar people from a certain population. In Experiment 1, the targets were famous Dutch males. The kind of instructions that are given to participants about the target pool is important before they attempt naming the composites. With a larger target pool, if participants are simply told the composites represent celebrities, as was the case in Frowd et al. (2005a, 2005b), the task can be too difficult, due to a degree of error being present in all composites, making them more difficult to name. In Experiment 1, participants were informed that the composites represent famous Dutch males who they might have seen in the Dutch television program “Opsporing Verzocht” (Leijtens, 2016).

This is a more specific instruction; however, the targets were from four different occupational fields (singer, actor, TV presenter, lawyer), keeping the target pool rather wide.

It can also be problematic if all the targets are not as current as the others. Before participants attempted naming In Experiment 1, they were informed that the composites represented international level footballers who were either current or former players in the UK Premier League. It was also deemed necessary to state that some players might have retired (fairly recently) or moved to play football at a lower level, because otherwise participants may have thought of players who had retired a long time ago, such as Paul Gascoigne. Yet, this still occurred with some participants who gave names of older players, for example David Beckham. These instructions to participants had to be included since otherwise some of the targets - Ryan Giggs, Steven Gerrard and Frank Lampard for example - would not have been considered by the participants. Although the targets were highly familiar to the participants and were identified at the rate of 95.4%, the fact that not all players were regularly seen on a football pitch is likely to have posed a problem in terms of who the participants included in their perceived target pool, and it would have probably been better to select targets who were all current players. This was the case in Fodarella et al. (2021), Experiment 2, in which the average correct naming rate for current football players was higher combined for EvoFIT and PRO-fit (13.8%), but for PRO-fit only this was significantly worse at 5%. This result is in line with the Experiment 2 in this thesis, that also used a feature system, which suggests that perhaps the two methods of composite construction in this experiment were also not effective enough. It is worth noting, though, that Kuivaniemi-Smith and Frowd (unpublished: see Appendix B) also used international level footballers and the accurate naming rate was much higher (13.9%) than in Experiment 2. The naming of composites was carried out when most of the players were still playing in the Premier League, and thus, they could have been deemed to be a more current target set.

In all the experiments in this thesis, foil images were used among the composite naming packs. The aim of them was to avoid a process of elimination from a known target pool (Frowd, et al., 2015). The foils were unknown faces of the same demography as the composites and matched the technique of the composites. For example, sketched images were used with sketch composites and PRO-fit images with PRO-fit composites, so that they would blend in and not be distinctive. Foil images have been found to suppress correct naming, although with a small effect size, and also to increase inaccurate naming of composites (Frowd et al., 2015). Given that the target pools discussed above appeared to be somewhat problematic, the presence of foils would not have made the naming task easier for participants.

4.2.2 Naming increased with a more specific target pool

In Experiment 3, footballer targets were again used; however, they were selected from one team playing below the UK Premier League. This choice was partly for logistic reasons, to enable efficient data collection for composite naming at a local football stadium, and also to increase the likelihood of gaining useful naming data for analyses. It is perhaps common sense to think that when there is only one team of players to consider instead of several, the likelihood of recognising someone from a composite increases. In a real investigation scenario, a composite aims to bear enough resemblance to the offender so that it can trigger someone's recognition of a person who is more or less familiar to them. This could be someone who is known well, in which case other factors such as a potential criminal history or the location of the crime might contribute to realising the composite could be a person they know.

In addition to average correct naming data, it is useful to look at the number of times a target is named based on a composite. In Experiment 3 one can see, that composites from other conditions, apart from 5 second no reference condition, were named most of the time by at least one participant. In a real investigation, any leads given to the police should be investigated and even one name could generate a lead to identify the offender. With average correct naming, we investigate how effective a given method is in general, but in reality, a composite that bears enough resemblance, or captures a distinctive feature, may be a successful outcome.

A large body of research indicates that faces that deviate from average are more recognisable (e.g., Shapiro & Penrod, 1986). They could be more unique looking overall, or have a particular feature that stands out, for example small dark eyes located close together. Hairstyles can also often be very distinctive; however, this is not a permanent feature as it can be changed and is therefore not as reliable as internal features such as eyes, nose and mouth. Distinctive traits can be particularly helpful for recognition of targets in smaller target pools. The overall distinctiveness of the target faces was not evaluated in the experiments of this thesis, unlike in for example Frowd et al. (2004, 2005a & 2005b) studies. Whether a face is highly distinctive or not can tell us how the different composite systems perform. In Frowd et al. (2004), E-FIT performed significantly better in constructing targets of low to medium-level of distinctiveness than highly-distinctive targets, and much better than EvoFIT. However, these systems performed equally for highly distinctive targets, and the performance of EvoFIT improved significantly for them compared to medium level distinctiveness. In Frowd et al. (2005a) on the other hand, all systems performed best with targets of higher distinctiveness. These systems were E-FIT, PRO-fit, Sketch, Photofit and EvoFIT. It seems that more information on how effectively sketches and PRO-fit composites are constructed for different types of faces would have been helpful to know, and for consistency of the targets faces of equal distinctiveness level could be selected before the experiment.

Experiments 4 and 5 also had more specific target pools, both of which included targets from a TV soap programme. The accurate naming rate for these experiments appears to be somewhat in line with naming of EvoFIT composites, but much better than feature systems, with similar targets (Fodarella, et al., 2021; Frowd, et al., 2015). Sixty second encoding duration was used in both experiments, and as found in Experiments 1-3, the fact that reference materials were not used, did not affect the composite outcome negatively (cf. using reference materials).

4.3 The role of target stimuli in composite construction

It can be debated what type of study stimuli would mimic real life in the best way. No significant difference has been found in the two most commonly used stimuli in facial composites studies, still images and video clips (see meta-analyses Meissner & Brigham, 2001; Shapiro & Penrod, 1986). This did not appear to affect the outcome of the composites. It could be argued that these are still not realistic enough and larger effect sizes for correct details in witness interviews have been found for staged events (as compared to video films) and if the interviewees actively participated in the event (Kohnken et al. 1999); however, when a limited number of factors are controlled in a study of relatively small sample size, using still images and video clips as stimuli are seen as acceptable. Staging a robbery, for example, involves relatively more resources and is more difficult logistically, which this thesis simply did not have an opportunity to implement.

As was discussed in the introduction chapter, certain views of the face offer benefit for recognition. In particular, a three-quarter view has been found to lead to better recognition compared to a frontal view (e.g., Bruce & Young, 1998; O'Toole et al., 1998). In the experiments carried out for this thesis, a frontal view of the face was drawn due to the stimuli being front facing images, hence no alternative view could have been formed in the participant's memory. A still image is one representation of the face. When a facial composite of this representation is constructed, how familiar the person viewing the composite is with the face it is attempting to depict is likely to have an impact on their ability to recognise them from the composite. Armann et al., (2016) tested memory for specific images and faces by using unfamiliar versus familiar faces. They found that participants were better able to remember a specific image of a familiar face than an unfamiliar face when different pictures were used at study and test. In contrast for familiar faces, the opposite occurred. If the person attempting to name a composite as someone who they are more vaguely familiar with, and possibly have not seen for a few years, they are likely to be dependent on how they looked when they saw them more regularly. For example, a footballer, whose hairstyle was different to a couple of years ago, as opposed to a more recent hairstyle which they have not seen. This may lead the viewer of the composite to exclude it as representing this footballer.

In Experiment 2, some composite evaluators commented on the target's picture "not really looking like them", which indicates that they were relying on a mental image different from the target picture. The level of the evaluator's familiarity of each target face was not recorded in the naming task, and thus this could not be included in the analysis, but this is something to consider for future studies as it could tell us more about how differing familiarity level affects recognition of the composites. Some faces from a target pool could have been seen in another context, for example a footballer in a current TV programme, and this could have acted as repetition priming, which has been found to enhance recognising celebrities faster by showing a picture to participants prior to test compared to unprimed pictures (e.g., Bruce et al., 1994; Bruce & Valentine, 1985). In real life, some people will be less familiar with the person being seen, while to others they might be very close (i.e., much more familiar). The challenge here is for the right people to be exposed to the composite, so that they have a chance of recognising the face. People who are less familiar with a suspect, having only seen them a couple of times, for example, might not consider the composite as someone they know, especially if the external features such as hair contain error or are different than what they remember. In a laboratory study more factors are introduced to the composite evaluation when people already have a varied familiarity level with the target pool, and more so if the criterion of the pool is wider than one containing a more specific group of people.

In Experiment 1, video clips were used that offered multiple angles of the target faces. The video clips were extracts from a Dutch talk show and, in addition to the target, they contained some of the audience in the background. The researcher remembers a couple of composite constructor-participants saying that they did not only look at the target face during the target view, but also the people in the background. This could have interfered with their memory of the target. They had not been instructed to only look at the target who was talking. Both seeing a person's face and hearing their voice is likely to bring advantages (e.g., Campanella & Belin, 2007); however, since the targets spoke Dutch, the participants did not understand what they were saying, and thus, no meaningful words could be associated with the targets.

It would be interesting to test in future studies whether encoding different orientations of the face, and then depicting the face in, for example, three-quarter view, or possibly in more than one view, could lead to better memory of the encoded face and therefore better identification of the composite.

4.4 Investigating the impact of context on composites

Chapter 3 explored the use of an effective mnemonic of the Cognitive Interview, mental reinstatement of context, in composite construction. The usual way to facilitate this technique is to ask a witness to think back to the event when they saw the offender and to recreate the environment, along with any sensorial cues connected with the event and also the witness's internal state of mind and emotions, as this is likely to enhance memory of the target face. A more detailed mental reinstatement of context, here called Detailed CR, involves the witness describing the context verbally. Fodarella et al. (2021) found a reliable effect of Detailed CR when composites were constructed with EvoFIT and PRO-fit systems. However, in one of their experiments, only EvoFIT composites were improved and not PRO-fit composites. Similar results were found in Experiment 5 of this thesis. The Detailed CR only improved standard sketch composites (cf. Minimal CR "think back MRC") but not PRO-fit composites. Fodarella et al. (2021) suggested that incidental encoding of the environment (as participants were not specifically instructed to take note of the environment) might lead to inconsistent findings and the outcome of composites will depend partly on this. Events are encoded differently by individuals, and while some may pay a lot of attention to the environment, some may not, which in a small sample may be revealed as inconsistency in results.

In Experiment 4, which manipulated context in three ways (Detailed CR, Minimal CR and Physical CR), target encoding occurred at a university restaurant, one which had been recently refurbished. Some participants reported having been to this area before, and thus there was an element of familiarity. In hindsight, these participants should probably not have been included, as was the case in Fodarella et al. (2021), since their presence in the sample may have impacted upon the effectiveness of the environmental cues (Hockley, 2008). A familiar environment has association with different events through personal experiences, which may be confused with each other, and this can potentially elicit memories from different events at retrieval (Hockley, 2008). Detailed CR led to impairment of sketch composites and were significantly worse than composites in the other two context conditions. This potential methodological issue was overcome in Experiment 5, in which the target encoding environment was chosen to be unfamiliar to witnesses in the study (undergraduate students).

Facial feature catalogues were used with sketching in Experiment 5. Encoding duration was 60 seconds in all conditions. No main effect of composite construction was found; however, it is clear that sketches were named much better with detailed context reinstatement than the other methods, and the interaction approached significance. This condition appears to have been the only one benefitting from detailed CR. It is possible that using reference materials after the initial sketch had been developed offered the participants’ useful cues for further retrieval. The benefit of reference materials was not evident after 60 second encoding in Experiment 3 compared to standard sketch; however, in Experiment 5 recall is likely to have improved by the detailed context reinstatement, leading to more identifiable composites.

It is worth considering how the mental reinstatement of context mnemonic could be potentially enhanced to further improve sketch composites. As discussed in the literature review of Chapter 3, Dando et al. (2009) found that a sketch version of the MRC elicited information as accurately as the “normal MRC” and with fewer confabulated items. Sketch MRC could be tested with sketch composites in connection to target encoding duration. There was an indication of the detailed MRC enhancing a weaker memory and an interesting question is whether sketch MRC could enhance it further. Often, when a composite sketch is constructed, there are moments in the interview when the witness is waiting for the sketch artist to make changes to the sketch—for example without being involved in the process as such (in an attempt to prevent interference of the face “under construction” with the mental image of the target). Witnesses may sometimes feel redundant in the process, despite the artist’s reassurance and rapport, and perhaps asking the witness to sketch the contextual information would make them feel more involved with the process. This may encourage further use of drawing from their memory, which is a method that the artist might suggest at the beginning of the interview. When words fail to describe, drawing could work better. Since repeated recall is also recommended (e.g., Fisher & Geiselman, 1992), using the sketch MRC could be used twice. First, the artist could draw the contextual information the witness describes, and after this the witness could work on this sketched context, which could make the retrieval process deeper and more effective still.

4.5 Factors to consider including for future research

4.5.1 More diversity in the target selection?

Most facial composite studies have used white Caucasoid target faces and often male (see Frowd et al., 2015), possibly due to the consistency of the target pool and therefore reducing other variables such as the cross-race effect (Malpass & Kravitz, 1969), should a more diverse set have been used. In Lech and Johnston (2011) two target faces were used (Black and White targets) and composites were constructed by four White participants and four Black participants. All composite evaluators, however, were white. There was an indication of a cross race effect, which with the small sample size could have skewed the results of the experiment. Laughery et al. (1977) also used black males as targets in one of their experiments and found that since most participant-witnesses were white, composites of the white targets were better likenesses than those of the black targets, giving more evidence to the cross-race effect. Kuivaniemi-Smith et al. (2014) also used a variety of targets who were from white, black and Asian (Indian) backgrounds, in a study which investigated remote interviewing for composite construction. This factor was not included in the analyses; thus, it cannot be said whether a cross-race effect was present or not. Ethnicity of composite constructors and evaluators varied in all the experiments in this thesis (as is the case in most composite studies) and this was not a controlled factor included in the analysis. Overall, it was found that remote interviewing was a feasible method.

More diverse target pools are desired for future research since this would reflect real life situations better in our increasingly diverse societies. Preferably a larger sample size than what is typical for composite studies would be recommended. Diversity is likely to add to the challenge of achieving a likeness of an offender with a composite in some cases. It is imperative that the ancestral origin is depicted accurately in a composite for it to be recognisable. Imagine a situation where, for example, an East Asian offender resembled a white Caucasian person too much, or vice versa. Fundamentally of course, the composite practitioner must follow the instructions of the witness, who may have perceived the offender's race wrongly. However, if the verbal description of the witness contradicts the appearance of the composite, it is the composite practitioner's responsibility to attempt to clarify, without any suggestive or leading questions from the witness, whether the composite represents the correct ancestral origin. An offender could of course be of mixed-race origin. This may present even more challenges for achieving good likeness, as it also does with facial reconstructions, in which resemblance depends on the details of the skull (see Wilkinson, 2004). A wide variety of answers on an ancestral origin of a mixed-raced clay facial reconstruction given in a survey highlights this difficulty (Robinson, 2010). Having a skull as a reference for facial features is based on science more than eyewitness memory also since DNA information may be available. However, subtle variation in appearance based on where an individual's geographical roots are, is still difficult, if not impossible, to depict. How much this issue affects the depiction's identifiability in composites is, to the author's knowledge, not known, but it is argued that this is not as fundamental as presenting the main traits of the face, including race/ancestral origin, correctly. A composite with more subtle errors in these aspects is more likely to be effective (someone recognising it as a person they know) than a composite with fundamental errors (e.g., wrong race, age or sex).

A facial composite might be in some cases aided by DNA. This is rare, however, since in relatively few crimes, no physical evidence of the offender is left behind (Peterson et al., 2010), and the percentage of these investigations requiring a facial composite is even lower. Parabon Nanolabs is a company which has developed a technique that can predict some facial features from a subject's DNA and create a facial image by reading tens of thousands of genetic variants ("genotypes") from a DNA sample (Parabon Nanolabs, n.d.). This approach would seem to help difficult cases such as depicting mixed-race individuals; however, according to Parabon Nanolabs, a forensic artist is still required to conduct a cognitive interview and produce a composite from a witness's description. A resulting composite would therefore be a combination of science and eyewitness memory. Since this scenario is highly unlikely in most composite cases, a composite practitioner is in a unique position when depicting a face from a witness's memory and description as accurately as possible.

It is also good to remember, that DNA phenotyping technology does not create a unique face representing an individual, but rather actively generates a suspect by clustering: producing a composite face of a collective, that can draw focus on some groups of people in the population and exclude others (M'charek, 2020). The incompleteness of a composite face aims to generate attention in the general public to the crime and invites them to specify it (M'charek, 2020). And this is in the hands – or rather – minds of the viewers: the connections they make by potentially recognising the face as someone.

A witness may describe an offender as looking "Eastern European". In such cases, care needs to be taken for the composite practitioner not to project their own perception of how this origin would be rendered into the composite image. If the composite system, such as standard sketch, relies mostly on recall, the task of achieving the “right look” is likely to be even more challenging than when reference materials are used as an aid for recognition. Or, if a system based more on recognition (e.g., EvoFIT) is used. EvoFIT for example, contains databases of White, Black, Asian, Eastern European, Chinese, Hispanic and various mixed-race combinations of both male and female faces of different ages (Frowd et al., 2012). For reference materials to be useful in sketching in this situation, a wide variety of different races should be available for a witness to view. Steinberg’s Ethnicities Catalog (2010) contains 18 chapters depicting thirteen separate ethnic groups, which is oriented to the population of the USA with representation of ethnicities such as Mexican, Puerto Rican, Cuban, Salvadoran, Dominican, Guatemalan, Columbian, Ecuadorian and Peruvian. This information is likely to help when a witness is attempting to refine facial features to resemble the individual more and do not have the vocabulary to describe how someone of a particular ethnicity looks like.

4.5.2 Individual differences in composite constructors

It is clear that differences in individuals are a factor that influences how effective a composite will be. This has been demonstrated with Photofit composites: participants, who produced better composites, did so consistently in two experiments (Ellis et al., 1975). Photofit is widely regarded as having limitations as a composite system (Frowd et al., 2005b); however, differences between individuals in composite construction is likely to transfer across systems. Some people simply are better at processing and recognising faces (e.g., Russell et al., 2009), as their memory for faces and detail is better (e.g., Fysh, 2018), or their verbal ability is superior on average (e.g., Tyler et al., 2023), and other people fall behind these standards, either to or below the average level.

Motivation is also an individual attribute that a composite constructor possesses and is a level of this varies between witnesses. A majority of composite studies employ university student (and staff) members. For Psychology undergraduate students, it is strongly encouraged that they participate in research projects conducted by other students, and they receive course credits for this. Anecdotally, it can be said that most student participants make at least some effort to achieve likeness in a composite. This is indicated by the comparable average times construction takes. There are also some students who are very motivated and genuinely interested in the topic of the experiment. However, some students do not appear to make an effort during cognitive interview and composite construction, which most likely affects how effective (or not) the composite will be. The latter two examples might be outliers in the study, and composites may perform according to the participants' efforts, with composites that either are highly identifiable or identified only by a couple of composite evaluators, if any. The success of a composite naturally depends on many other factors, however.

It has been found that motivation is correlated with accurate information. In Paulo et al. (2015), student participants watched a short video of a bank robbery and were interviewed using either an enhanced cognitive interview (ECI) or a structured interview (SI) 48 hours later. After the interview, participants were asked to rate their motivation on a seven-point Likert scale (1 – very low; 2 – low; 3 – slightly low; 4 – moderate; 5 – slightly high; 6 – high; and 7 – very high). No difference was found between the interview conditions; however, participants' own perception of their motivation during the interview correlated with report accuracy (correct recall proportion). The ECI elicited more correct information overall, without increasing errors, compared to the SI. In this study, it was suggested that participants rated their motivation from moderate level upwards due to the inclusion or greeting and rapport in the interview. In composite construction, particularly sketches, there are moments when an artist is focusing on drawing the face and the witness is required to wait until they are shown the sketch again to review (after alterations requested by the witness). This is when there is a risk for a witness's motivation to lapse, although this is more likely in a mock case where no real criminal is being sought. A witness is likely to have high motivation in helping to catch an offender who committed a crime against them or someone else. Fisher and Geiselman (1992) advise re-establishing rapport with a witness if motivation is detected to be decreasing. This instruction was followed in experiments of this thesis. It is certainly worth considering including motivation ratings in future studies to investigate how this might affect the effectiveness of composites.

Studies that use participants from a wider pool of demographics and backgrounds for composite construction, may benefit from a higher motivation to the task in general. There is some indication on this. In Kuivaniemi-Smith & Frowd (2013) Experiment 2, composite constructors were recruited from the community living nearby the researcher, and the study was conducted in various places such as a local library or their home. Encoding and composite construction were carried out in different places to avoid physical reinstatement of context. The correct naming rate for this study was high, and much higher than Experiment 1; however, the target pool differed so that in Experiment 1 targets were international level footballers and in Experiment 2, they were Coronation Street characters, and this is likely to have affected the difference also. Motivation of the constructors is argued to have been high since the participants made more effort to take part in the study than what was required from student participants. In hindsight, constructing composites at participants' houses might have introduced additional factors in the study. Encoding occurred elsewhere than in a home environment, however. If this had been the other way round, the results might have been different due to a familiarity effect. Mental reinstatement of context was facilitated in the usual way to a witness interview ("think back to the event"), and not in a detailed manner, thus, the effect would possible have been less impactful as indicated in the findings of Experiment 4.

Similarly, in Experiment 4 of Fodarella et al. (2021), composite constructors were local residents for a local town, thus their motivation to the task is argued to have been elevated compared to student participants in general. The overall correct naming rate for composites of EastEnders characters was much higher than in their previous experiments for composites.

4.5.3 One or more artists?

It could be argued that using more artists in research studies would add ecological value to the research. While there are guidelines on the interview process and procedures concerning eyewitness memory (e.g., ACPO, 2009; Richardson et al., 2010), there are no international guidelines on the sketching technique itself (Davies & Valentine, 2006) and sketch artists find their own ways through experience, and thus it would be difficult to fairly compare composites different artists produce due to so many uncontrolled variables. This is also tricky since the witnesses' memory and retrieval abilities are always unique and if one artist interviews a witness who has a very clear mental image of an offender as opposed to another artist interviewing a witness with a weak memory of the face, comparison between the two artists might not fair or reliable, which is likely to make artists apprehensive for such comparisons. They can only help the retrieval to a point yet have no control over the witness's storage of the memory as such. Some comparison in artistic skills can be found from the research literature. Laughery (1977) conducted three facial composite experiments with a large sample size and found individual differences in artists but not in Identi-kit technicians. This indicates that due to the more flexible nature of the sketching technique, it is more dependent on individual skills than a more rigid system. It was somewhat impossible to conduct a research project with a larger sample size for this thesis and with a smaller sample these individual artist/participant differences may have become more apparent. It was therefore seen fairer for one artist to conduct all composite constructions in all the experiments to increase the power of the experiment, which is admittedly still not a perfect solution. Future research could consider using more artists in an experiment if the sample size is increased.

4.6 Evaluation methods – are they reliable?

Since 2005, facial composite research has had a "gold standard" for conducting experiments (Frowd et al., 2005a, 2005b). One criterion in this standard is that participants who construct the composites are unfamiliar to the target faces they encode, to reflect real life. Another criterion is that participants attempting to name the composites are familiar with the target pool that was used for composite construction, so that they have an appropriate memory for those faces to match the composites to, should one look familiar to them. This is said to mimic the real crime scenario because the aim of the composite is for someone, whether within the police or a member of the public, to recognise it as someone they are familiar with. Naming is often considered as the main method of evaluation for this reason. As discussed earlier in this chapter, evaluation by naming is not without its problems and careful consideration is required in the design for it to be as effective as possible and to measure what is intended.

In Experiment 5, the naming task was conducted either face to face, as in other experiments, or remotely via a Skype call. The remote method had not been used by the researcher before for naming, and it is possible that the methodology could have led to some differences in the results compared to conducting it in person. The composites were naturally presented differently, and while in the manual condition participants were fully in control of the pace of looking through the paper images, in the Skype condition they often needed to be prompted to verbally state when were ready to move to another image. The composites were presented to participants by the researcher sharing her screen, which could have led to the images being focused on in a different way. For the methodology to be closer to the face-to-face method, participants could be given access to the composite images so that they could look through them more independently than via screen sharing. Alternatively, if the order of the composites was randomised after each participant, all the composites could be presented in one word document, showing one composite on each A4 page.

Other evaluation methods include a sorting task, identification parade and a likeness rating task. Likeness ratings have been part of composite research since the first study using the Photofit system (Ellis et al., 1975), and thus have a long history in composite evaluation, and should rightly still be valued as a means for assessing how effectively composites can be constructed with various techniques, and not purely as a supplement to naming.

In all the experiments in this thesis, composite raters were instructed to assess the overall likeness of the composites while comparing them to the target faces, so that holistic face processing is encouraged. The rating process contains a varying degree of subjectivity because each rater will evaluate the faces differently. This may include emphasis of a distinctive feature (i.e., if the feature is well captured in the composites, the evaluator might give a better rating for this reason), or an evaluator of a different race to the composite might give more emphasis to different features than an evaluator of the same race as the composite. Cultural differences, for example, have been found to affect faces being processed differently by western Caucasian participants and east Asian participants (See Wang et al., 2020).

In Experiments 3 and 5, likeness rating commenced via an online survey. This layout presented the composites vertically one below another and each composite had the target image next to them. Even though all images were visible together, this presentation format could have encouraged participants to evaluate the composites one by one, more so than in other experiments, where the composite images were laid on the table side by side with the target image above them. And if the composite plus target pairs were viewed on a smaller screen (e.g., mobile phone), or zoomed in, they would have been viewed more or less individually. Evaluating all composites together with the target allows comparison between the composites themselves and with each composite-target pair simultaneously. In addition, due to participants being instructed to mentally rank the composites from best to worst while rating the composites, the task would have encouraged comparison between the composites further. This was seen to help the participants in their rating process to be more accurate and is likened to a within-subjects experiment, which has more experimental power (Jackson, 2023). Jackson (2023) used this same technique in several of her experiments.

In most composite studies that have included likeness ratings, ratings have been given by evaluating one composite and its target sequentially, which does not allow immediate comparison between composites. The raters were required to be unfamiliar with the targets in all the experiments in this thesis, as familiar faces may be rated differently (more harshly) (Frowd, 2021). With familiar faces, it is argued that assessing all the composites simultaneously would be more of a problem due to this. However, since unfamiliar faces do not have the same issue, it is argued that this potential bias, highlighted by comparison of different composites, is not present in the same way.

4.7 Limitations of laboratory research

Cognitive Psychology has contributed enormously to understanding human behaviour. However, it also has limitations. One concern is that laboratory research lacks ecological validity—how it reflects real life situations (e.g., Banaji & Crowder, 1989; Cohen & Conway, 2007; Conway, 1993; Eysenck & Keane, 2015), especially when the witnessed event has been highly arousing or violent (Yuille, 1993; Yuille & Cutshall, 1986). Banaji and Crowder (1989) believe this is due to the inability to control unique situations, how effectively information is encoded and what occurs between the encoding and retrieval stages for each individual. And Cohen and Conway (2007) suggest that in the real world a researcher has no control over the initial learning phase (for example, with faces). They argue that there are several factors affecting the memory which cannot be controlled: the degree of attention paid to the face, the number and duration of encounters and the quality of these encounters, and the number of especially similar faces encountered during the intervening period.

More traditional researchers, who endorse controlled research, on the other hand argue that everyday memory research fails to generalise because uncontrolled factors are allowed to vary freely (Cohen & Conway, 2007). Laboratory based facial composite research minimises or eliminates the effects of different variables such as varying lighting, visibility and distance of the target face, by allowing participants to focus on the study stimulus while being aware of having to recall the face later (intentional encoding) and construct a facial composite of the target. In everyday memory, the remembered information has often been learned incidentally rather than intentionally (Cohen & Conway, 2007). Kuivaniemi-Smith et al. (2014) takes this into account in one of her experiments investigating remote interviewing; participants viewed the study stimulus while not being aware of the purpose of the task at that stage. Only afterwards was it explained that participants needed to describe the person they saw from their memory and construct a facial composite the next day. The experiments of this thesis all used intentional encoding for consistency between the experiments, and because much of the facial composite research using the “gold standard” procedure, to more closely mimic real life (Frowd et al., 2005a), has the same design (see meta-analysis by Frowd et al., 2015).

The sample size of composite construction research is often rather small, but with at least ten participants per condition and ten targets (Frowd, 2015), it has sufficient power to detect a medium-to-large effect size for the purpose of meta-analysis (Fodarella et al., 2021). Some attributes of the real-life scenario are applied to these experiments: once the study stimulus has been encoded by participants, there are no control measures in place as to what occurs in a participant's life between encoding and later retrieval, and the same would apply to witnesses of crime. No instructions are given to participants to memorise the face in a particular way or keep remembering it during the intervening period, and therefore this stage is based on the participants' unique memory capacity and processes in a free manner. Some participants may report that they have tried to focus on their mental image many times, but others may have even forgotten about the task until the composite construction stage. How this affects composite quality is impossible to comment on without a more controlled study design that would focus on individual capabilities. On the other hand, everyday memory research does take individual differences of people such as age, culture, personality, educational background into account more than many laboratory studies that recruit mostly young students for convenience (Cohen & Conway, 2007). However, a more diverse sample from the wider community was recruited in Experiment 1 of this thesis due to inconvenient timing of the composite construction phase in the university settings.

Kvavilashvili and Ellis (2004) have critically evaluated the ecological validity of research and their findings indicate that it is not an exclusive property of everyday memory research, and in fact a carefully designed laboratory study may score higher in both representativeness and generalisability. For example, the EvoFIT system (Frowd et al., 2004) have been researched rigorously in laboratory studies for years and the performance of this system has improved from constructing less identifiable composites than feature systems (e.g., Frowd et al., 2005a; 2005b) to being the benchmark in facial composite construction (e.g., Frowd et al., 2011; 2015; 2019). It has also been tested in field studies (e.g., Frowd et al., 2011), which have demonstrated the effectiveness of the system in criminal investigations, constructing successful composites that have led to identifying an offender in serious crimes (e.g., Frowd et al., 2012). It is worth mentioning that there will always be limitations when studying witness memory and there are factors that cannot be controlled by the researcher. It has been acknowledged that in some instances, it can be impossible to know whether the original memory has been altered by post-event information or whether a second memory has been formed that competes with the original memory at the time of test (Loftus & Loftus, 1980).

4.8 Can the naming of composites be improved after they have been constructed?

Most facial composites include an inherent degree of error, despite the system used to construct them. There are some postproduction methods that have been applied to composites to explore if they can be made more recognisable after being constructed. Frowd et al. (2007) found a negative caricature effect (de-emphasising features) for composites constructed by E-FIT, PRO-fit and sketch. Volunteers, who were aware of the celebrity composites' identity, were asked to make the faces as identifiable as the target as possible. The caricaturing effect was applied at different fixed stages. It was suggested by the researchers that the superiority of the negative caricature was due to it reducing some error often present in composites since they are not portraits and predicted that the anti-caricature effect should be most effective for composites of poor likeness. Morphed composites were included in another experiment of similar design by Frowd et al. (2007). They were found to have an opposite caricaturing effect: the composites were seen as better likenesses with a positive caricature effect (emphasising features). Since morphed composites tend to reduce error from individual composites, these were seen to benefit from caricaturing by emphasising the features (in the whole face) (Frowd, et a., 2007). Another experiment in Frowd et al. (2007) measured the effectiveness of a moving caricature. Participants' preferences in the caricature level varied widely from moderate positive caricature to a slight negative caricature. It was found that composite naming increased by as much as 50% when the composite was seen to change in small steps (5%) from -50% to +50% caricature. This was effective for all composite systems mentioned above but those composites that were initially named the lowest had the most benefit from this procedure -they were named up to 10 times better. Dynamic or positive caricature could potentially be a very good method to increase the identifiability of sketched composites and should be explored in future research.

Another post-production method for composites is viewing the composite side-on, so that the face appears long and thin (e.g., Frowd et al., 2013). Frowd et al. (2013) tested the impact of stretching the composite either vertically or horizontally so that the configural information of the image changes and found that vertical stretch was beneficial for composite naming. Stretch also led to significantly higher naming than veridical (unchanged) image; however, this was in the cued naming. Viewing the composite from the side or from below mimics the physical stretching by causing a "perceptual stretch", which considers the practicalities of showing composite faces to the public for example (Frowd, et al., 2013). In another experiment, Frowd et al. (2013) applied vertical instead of horizontal stretch, since the latter has been found to interfere in recognition of photographs (Hole et al., 2002) and also because of the results of their previous experiment. Perceptual stretch/viewing the image sideways led to significantly higher naming than did viewing normal, front on faces. In fact, the researcher recalls an interesting event when constructing a composite sketch with a participant on a Skype video call as part of her research project on a Forensic Art course. While the participant was viewing the sketch held up to the web camera by the researcher, the sketch moved changing the proportions slightly and the participants asked her to tilt the composite to see this again and identified this position as a better likeness to the subject depicted from her memory. Sideways view of sketches has been explored by Kuivaniemi-Smith and Frowd (2013) to investigate whether this could improve the naming of sketches. This generated some mixed results that were not enough to conclude whether sideways view is helpful or not. This issue would be worth exploring further.

4.9 Further ways to improve sketches

The Self-Administered Interview (SAI) (Gabbert et al., 2009) aims to reduce degrading of recall by instructing witnesses to recall an event, on their own, as soon as possible after a crime. This could be especially helpful for sketch composites that require fine details of facial features to be recalled. Some evidence of this already exists. Robertshaw (2020) found that when participants wrote down their descriptions of a target face 3-4 hours after encoding, this resulted in more identifiable composites, than when written descriptions were not included in the process. Some witnesses make notes spontaneously without instructions after a crime, as they anticipate needing to recall information later. No data on whether this practice is helpful when a composite is required, exists, but based on previous research findings, it is likely to help rather than hinder the outcome of the composite. SAI, combined with constructing sketch composites sooner, perhaps remotely, would be predicted to improve sketches further. In current times, prompted by the Covid-19 pandemic, many services have moved online, and people expect to have a remote option. According to Kuivaniemi-Smith et al. (2014), sometimes participants find positives in remote composite construction compared to face-to-face. For example, they felt less under pressure. In general, participants found the remote interview to be a good alternative to in-person interview; however, they would still often prefer the latter. It is good to consider that the composites for that experiment were conducted in 2010 and life has increasingly moved towards online services since then, with improved technology. It can be fairly confidently said that attitudes towards this issue have changed, and remote interviews would probably not be seen as out of the ordinary anymore.

Holistic Cognitive Interview has not been found to have a clear benefit for sketches (e.g., Stops, 2012); however, there is indication that it works in conjunction with the whole-face view of PRO-fit facial features, but clashes with the isolated feature view when the composite is continued as a sketch (Kuivaniemi-Smith & Frowd, 2013). In another experiment, it was also found that H-CI benefitted composites that had been named at a lower level (cf. Cognitive Interview), while the benefit dampened down with the higher-level naming. No difference between standard sketch and PRO-fit whole-face/sketch was found. While PRO-fit/sketch combination was not found to be effective in Experiment 2, standard sketch led to reasonable, and even high naming levels in Experiments 4 and 5. It would be interesting to continue investigating the potential effects of H-CI on sketches. It is suggested that the initial sketch stage could benefit from this procedure.

5 CONCLUSIONS

There has been limited research on composite sketching, and this thesis started addressing some key questions to investigate how existing composite construction practices and novel approaches impact the effectiveness of composites. Except for the cognitive interview, sketch composite construction is not regulated as a technique. Some facial identification related rules are recommended to follow, however, which include not showing witnesses, or an artist, any pictures of the suspect(s) prior to composite construction, as this may interfere with witness's memory and influence the artist. Reference materials appear to be widely used among sketch artists, yet this key area of composite construction has remained largely unexplored, and artists have developed their technique through experience. This is of course important due to the practical nature of the work; however, since other composite systems from mechanical to computerised have received a lot of attention since mid-1970's, it was deemed necessary to start investigating what the role of recognition cues is in sketch construction.

Many crimes occur quickly, and witnesses might not get more than a fleeting glance of an offender. Therefore, an important question was to ask what happens to face processing and memory in these cases. In the knowledge of sketching being heavily recall oriented, and recall being more vulnerable to information loss as the time passes than recognition, it was seen as one of the key areas to explore how sketching performs after a short encoding duration. The findings indicate that reference materials improve sketches when encoding duration has been short (5 seconds). This is valuable to know, as some sketch artists do not use any reference materials in their procedure due to the potential concern of interfering with the witness's memory. Since reference materials have been found to have a neutral effect after longer encoding and not impairing composites compared to standard sketch, it is recommended that reference materials are utilised by artists. Witnesses should be guided and encouraged to use the reference materials available, even when they feel they might not need them, as the demonstrated confidence in eyewitnesses is not a reliable measure of the accuracy of their memory. More research is required in how to optimally use reference materials, but based on the results here, reference materials should be used after the initial sketch has been drawn and presented to the witness as a whole face. What is not known, is whether it is better to see features in a facial feature catalogue in isolation from other features, or as part of whole faces.

A novel approach combining the PRO-fit system (for selecting facial features) and sketching, did not achieve the same expected outcome as the facial feature catalogues, and this method was no better than standard sketching, even after short encoding. In this method, pictures of facial features were selected in a whole face context after free recall, and thus less detailed recall was required (cf. standard sketching). Combining different composite systems in one technique clearly requires more research and it is too early to advise which method is best. EvoFIT includes a database of sketched faces, which is likely to work better with sketching due to the similar visual elements. In fact, there is a software called SketchCop (SketchCop®, n.d.) available that uses sketched facial features instead of photographic (such as PRO-fit for example), but this system presents the features in isolation from other features, which is known to be less effective than whole face processing of the features (e.g., Skelton et al., 2015). Other sketch-like feature systems include Identi-Kit 7, the latest computerised version of this system (Identi-Kit Solutions, n.d.).

Directing the participants' attention to the encoding environment in a more detailed manner by asking them to describe contextual information verbally, led to some promising findings in Experiment 5 for sketches. This implies that recall was enhanced, and sketching was an adequate tool for constructing composites that received a high naming rate. It is not known how comprehensively cognitive interview, including mental reinstatement of context, is conducted by different sketch artists around the world, but it is included in the guidelines in the UK (ACPO, 2009) and USA (Richardson et al., 2010). Experiment 4 found that detailed context reinstatement led to worse composites than both minimal CR and physical CR. The encoding environment was somewhat familiar to some of the participants, which was suggested to have impacted the results. In Experiment 5, encoding occurred in an unfamiliar environment to all participants. If a crime has been experienced in an unfamiliar place by a victim, it is recommended that detailed context reinstatement is facilitated prior to free recall in the cognitive interview. If, however, a crime occurred in a familiar place, a usual way of instructing the victim to think about the context (but not elaborate by verbalising it) is advised. A familiar place is connected to the victim's previous experiences and facilitating detailed CR in these circumstances could be detrimental to the effectiveness of the composite, and thus is advised not to be used in such cases.

What this research has shown is that the sketching technique has a lot of potential to develop and improve further, and it should not be seen as a dying art to be moved aside by the computerised systems. Many methods, such as caricaturing the composite after it has been constructed, has been explored little with sketches, and is a worthwhile avenue for further research. Composite construction is a combination of different mental processes, especially perception, retention and retrieval. Understanding more about how these elements work together to complement each other will most likely lead to further improvement in capturing a sketched face from memory. For example, further improvement of recall, which is acknowledged to be more difficult than recognition, could also be harnessed as an aid to recognition and not only the other way around. The Self-Administered Interview, for example, could preserve detailed recall for facial features. This would potentially help in the creation of the initial sketch from witness's recall, which appears to be an important stage in sketch composite construction.

A composite aims for someone familiar with the face to recognise it. Since the circumstances are varied in real life: one suspect might be a known reoffender, who the police officers are familiar with, and this is when the target pool to recognise the suspect is narrowed compared to a suspect who is not known to the police before and whose identification relies on the members of public. This suspect may not even be considered as a possibility in being represented in a composite by the people who know him or her, and thus, a chance of identifying the composite is slimmer than in the first mentioned case. Anyone viewing the composite image should be aware that they are not looking at an exact depiction of an individual, but rather an approximation that aims to capture as many similarities with the individual as possible. As a forensic examiner Jackson (2004) states: "Does the drawing need to look exactly like the perpetrator to be effective? No, it does not. The likeness should be as accurate as possible, but a general or close likeness will in many cases stimulate recognition on the part of viewers." Jackson (2004) suggests that while it is commonly believed that highly detailed or photographic images are more effective to recognise a person from them, it is possible that these images narrow the scope in the viewer’s mind, who may expect the photographic image to have a perfect likeness to someone they know or have seen, to consider them as a suspect.

6 APPENDICES

6.1 APPENDIX A – Ethical approval

THE UNIVERSITY OF WINCHESTER

RKE ETHICS PROFORMA – STAFF

SECTION 1

Project title/focus: Exploring the importance of reference materials (pictures of facial features) in a facial composite construction method when the target exposure time is short or medium length.

Principal Investigator: Heidi Kuivaniemi-Smith

INDIVIDUAL PROJECT APPROVAL

Does the research involve:	*Please place tick in each column as appropriate*		
	YES	**NO**	
Living human participants	✓		**If yes, please complete section 2**
Living animals other than those being observed in their natural habitat		✓	**If yes, please complete section 4**
Documentary material that is not already in the public domain		✓	**If yes, please complete section 5**
Handling sensitive materials, including human remains		✓	**If yes, please complete section 6**
Interventions in the natural or built environment		✓	**If yes, please complete section 7**

Please refer to the University RKE Ethics policy for clarification of any of the above terms

SECTION 2: RESEARCH ETHICS CHECKLIST FOR WORK INVOLVING HUMAN PARTICIPANTS

	Research that may need to be reviewed by NHS NRES Committee or an external Ethics Committee (if yes, please give brief details as an annex)	*Please place tick in each column as appropriate*	
		YES	**NO**
1	Will the study involve recruitment of patients or staff through the NHS or the use of NHS data or premises and/or equipment?		✓
2	Does the study involve participants age 16 or over who are unable to give informed consent? (eg people with learning disabilities: see Mental Capacity Act 2005).		✓
	All research that falls under the auspices MCA must be reviewed by NHS NRES.		
	Research that may need a full review:		
3	Does the research involve other vulnerable groups: children, those with cognitive impairment or those in unequal relationships e.g. your own students?		✓
4	Will the study require the co-operation of a gatekeeper for initial access to the groups or individuals to be recruited? (e.g. students at school, members of self-help group, residents of Nursing home?)		✓
5	Will it be necessary for participants to take part in the study without their knowledge and consent at the time? (eg covert observation of people in non-public places)?		✓
6	Will the study involve discussion of sensitive topics (eg sexual activity, drug use)?		✓
7	Are drugs, placebos or other substances (e.g. food substances, vitamins) to be administered to the study participants or will the study involve invasive, intrusive or potentially harmful procedures of any kind?		✓
8	Will tissue samples (including blood) be obtained from participants?		✓
9	Is pain or more than mild discomfort likely to result from the study?		✓
10	Could the study induce psychological stress or anxiety or cause harm or negative consequences beyond the risks encountered in normal life?		✓
11	Will the study involve prolonged or repetitive testing?		✓
12	Will the research involve administrative or secure data that requires permission from the appropriate authorities before use?		✓
13	Is there a possibility that the safety of the researcher may be in question (eg in international research: locally employed research assistants)?		✓
14	Does the research involve members of the public in a research capacity (participant	✓	

	research)?		
15	Will the research take place outside the UK?		✓
16	Will the research involve respondents to the internet or other visual/vocal methods where respondents may be identified?		✓
17	Will research involve the sharing of data or confidential information beyond the initial consent given?		✓
18	Will financial inducements (other than reasonable expenses and compensation for time) be offered to participants?		✓
19	Are there problems with the participants' right to remain anonymous?		✓
20	Is the right to withdraw from the study at any time withheld, or not made explicit?		✓
21	Does any part of the project breach any codes of practice for ethics in place within the organisation in which the research is taking place?		✓
22	Is any of the material used likely to cause offence to any of the participants?		✓
23	Is a contract* needed between the researcher and the participants?		✓

Please refer to the University RKE Ethics policy for clarification of the above terms

*** 'Contract' includes requirement for written consent**

SECTION 3 ETHICS APPROVAL FOR WORK INVOLVING HUMAN PARTICIPANTS

Project title: Improving the construction of sketched-based facial composites

Principal Investigator: Heidi Kuivaniemi-Smith

The project outline/proposal, with any relevant papers, should be attached. This should include:

1. Showing clearly aims, objectives, location, methodology
2. Where involving human participants, comment on criteria for selection/interview and sample size
3. If consent is required, who is to give it?
4. Outlining potential risks to the participants and how you will deal with this
5. Stating the information to be given to the participants (attach copies of letters or information sheets that will be given to the participants)
6. Noting the published code(s) of practice being followed
7. Whether the project involves any other disciplines or local ethics committees

8. What arrangements have been made to ensure the confidentiality of data collected, and compliance with the Data Protection Act
9. Whether payment to the participants will be made
10. Whether the project will receive financial support from outside the University
11. Whether participants, employers or ethics committees have placed restrictions on the publication of results
12. Any further points you may wish to make in justification of the proposed study

A summary of the issues and actions taken to address the ethics issues

If you have fully explored ethical issues in your proposal, including reference to guidelines to which   *you will adhere, you need only refer the scrutineer to your project proposal*

DECLARATION C

For completion by those required to fill in Sections 1, 2 & 3.

Declaration: I have read the University's policy on ethics related to Research and Knowledge Exchange and to the best of my knowledge and ability confirm that the ethical considerations noted have been assessed. I am aware of and understand University procedures on ethics in Research and Knowledge Exchange and Health and Safety. I understand that the ethical propriety of this project may be monitored by the Faculty RKE Committee.

I will abide by the expectations of the University RKE ethics policy, and by the ethical guidelines published by British Psychological Society.

Signature of Principal Investigator: Date: 22/9/2014

Countersigned for Faculty RKE ……… …………. Date: ……2 October 2014……

RKE ETHICS SCRUTINY FORM (work with human participants)

Project focus/title: Exploring the importance of reference materials (pictures of facial features) in a face composite construction method when the target exposure time is short or medium length.

Principal Investigator: Heidi Kuivaniemi-Smith

The following points must be assessed:

1	Potential conflicts of interest	N/a ☐ Met ☐ Partially met ☐ Not met ☐ Comment:
2	The implications of monetary or other inducements	N/a ☐ Met ☐ Partially met ☐ Not met ☐ Comment:
3	The nature of recruitment and participation if the project involves participants from vulnerable groups	N/a ☐ Met ☐ Partially met ☐ Not met ☐ Comment:
4	Procedures for providing explanations to participants, including the preparation of an appropriate information sheet	N/a ☐ Met ☐ Partially met ☐ Not met ☐ Comment:
5	Procedures for obtaining informed consent from participants (or where applicable their parents or guardians), including the preparation of written consent forms	N/a ☐ Met ☐ Partially met ☐ Not met ☐ Comment:
6	Possible discomfort, distress or inconvenience to participants and/or Researchers	N/a ☐ Met ☐ Partially met ☐ Not met ☐ Comment:
7	Procedures for respecting confidentiality and operating with data protection legislation	N/a ☐ Met ☐ Partially met ☐ Not met ☐ Comment:
8	Safety risks in accordance with the University's Risk Assessment Procedure and measures taken as appropriate to make them as low as reasonably practicable	N/a ☐ Met ☐ Partially met ☐ Not met ☐ Comment:
9	Has an appropriate risk assessment been carried out and attached to the proposal?	N/a ☐ Met ☐ Partially met ☐ Not met ☐

		Comment:
10	Other issues you may wish to raise	Please comment:

Scrutineer's recommendation:

☐ Appropriate action taken to maintain ethical standards – no further action necessary

☐ Further emendation necessary

☐ No approval

Reason for emendations or approval

Signature of scrutineer: ………………………………………………………………… Date: ………………………

Project proposal

Facial-composite images are used widely in police investigations to identify criminal suspects. These images are constructed by (eye) witnesses using a manual technique (artists) or by computerised methods (e.g. E-FIT, PRO-fit and EvoFIT software systems). Artists construct sketches by asking witnesses to describe the appearance of the face, select facial features (via pages of eyes, nose, mouth, etc.) and build up the face by hand using pencils or crayons. The current project involves a new procedure, which potentially improves the identifiability of the sketch composites.

The standard sketching method uses no reference materials (pages of facial features) as a recognition aid and an initial sketch of the whole face is created lightly based on the witness' descriptions. From thereon, the sketch is shown to the witness and he/she alters the sketch according to their mental image of the offender. This is one of the methods used to construct the composites in the current project. Established research (e.g. Tanaka & Farrah, 1993) shows that much better identification of features occurs when features are selected in the context of a complete face. While the standard sketching method allows this, the focus may still be heavily on individual features as the witness works on these one after another. For the alternative facial composite construction method, the constructors select features seen in the context of a complete face using standard PRO-fit composite software to allow people to select appropriate facial features. This may make the process easier for the witness providing a recognition aid from the start of the interview. Once features have been selected in this way using PRO-fit, the researcher (artist) will draw the face by hand.

Research also indicates that much more identifiable composites are produced when people construct the face after having been asked to think about the character of the target (e.g. Frowd et al., 2008, 2012, 2013). This so-called 'holistic' cognitive interview (H-CI) technique contrasts with the previous approach whereby people are asked to describe the target's facial features using a free-recall procedure (via a 'cognitive interview', CI).

Unfamiliar face recall and recognition is a challenging task, especially if the target was only seen for a few seconds and this needs exploring further in the context of composites. There is some evidence on the effects of exposure time in facial recognition. Shapiro and Penrod (1986) found that as the time spent viewing a target increases, so does the correct identification of the target. Ellis and Flin (1990) showed two groups of 10-year old schoolchildren a face either for 2 seconds or for 6 seconds and found an advantage of the memory strength for the latter group immediately after the target exposure (Deffenbacher, Bornstein, McGorthy, Penrod, 2008). The advantage was more evident after longer (2 and 7 days) of the encoding process (Deffenbacher et al., 2008).

The proposed design for face construction is between subjects with two experimental factors: face construction (sketching with no reference materials, and using PRO-fit for reference materials combined with sketching) and target exposure time (5 seconds and 60 seconds). The expectation is that sketched composites will be more identifiable: (a) with PRO-fit used for reference materials than without use of reference materials and (b) with 60 seconds compared to 5 second target exposure, and (c) with 5-second encoding and use of PRO-fit compared to 5-second encoding without use of PRO-fit (i.e. there will be a significant interaction). Participants will be randomly assigned, with equal sampling, to one of these four conditions.

There are two stages to the design. Composites would be first constructed by recruiting participants to see a target face and then (the following day) produce a sketched image of the face using one of the two construction conditions (as described in the above 2 x 2 design). In the second part, composites would be evaluated for quality by asking further participants (a) to name the composites spontaneously and (b) to judge the likeness of the composites in the presence of the target.

The design aims to parallel construction in the real world by recruiting participants who are unfamiliar with the target faces (as witnesses of crime do not normally know the identity of the criminal of which they construct a composite) while the main evaluation of composites is carried out by different participants who are familiar with the targets, and so are capable of naming the composite images (Likeness ratings will be carried out by participants who are unfamiliar with the targets.). To achieve this objective, the project plans to use target faces of footballer players who play in top teams in the UK. Thus, participants will be recruited on the basis of either not knowing the identity of football players (face construction and likeness ratings) or those that follow the game (composite naming). The plan is to use images taken from the Internet of these players. These images will be of good quality, non-offensive and front-facilitating; there will be ten different top-level players in total. These images will be printed out a limited number of times for the project (four) and discarded after use. (The PI would like to note that using publicly available images in this way, as well as the proposed design in general, has been accepted as part of ethical applications in the past.)

The research will be conducted in accordance with ethical guidelines by the governing body for Psychologists, the British Psychological Society. (The project does not include any other disciplines or local ethics committees.) As part of this, participants will be given an information sheet to read through before agreeing or not to take part in the research. For those who agree, meeting times would be arranged. Based on past research, and to give sufficient experimental power, the plan is to recruit 40 participants in total for face construction; about 48 for face naming (12 participants randomly assigned to each of four conditions) and about 24 for composite rating (within subjects with participants inspecting all 40 composites). As face construction is in two parts, these participants will be given the information (briefing) sheet to take away with them, which informs them of their right to withdraw between sessions. All participants will receive a verbal debriefing and then a debriefing sheet to take away with them. The debriefing sheets will also remind participants of their rights to withdraw. None of the participants are likely to be stressed as a result of the procedures (I have run several studies like this in the past, without any negative issues).

Participants who construct a composite will be offered a £5 incentive to complete construction of a composite or alternatively they receive course credits; for other participants, participation will be voluntary. All of this information is explained clearly on the briefing sheets. The project will not receive external funding.

Face construction will take place mainly at the University of Surrey (assisted by an on-going collaboration there) and partly at the University of Winchester. Recruitment will be mainly by email to students and staff at the Universities and via the Sona System. All potential participants will be adult, will read the relevant information (briefing) sheet initially and will be able to give verbal consent. Composite naming and likeness ratings will be collected at the above-mentioned universities.

Information sheets for briefing (to be read prior to making a decision about whether to participate or not in the research) and debriefing (given out at the end of each research session) are listed below. These sheets also indicate that data will be collected anonymously and the intention of the research output (publication and change of practitioner practice, if appropriate); no restrictions will be made on publishing of results. The research will record participant age and gender, if participants' consent. If participants' Christian names are recorded, for scheduling purposes, this information will be written on a separate sheet to the participants' data and will be discarded once the person's experimental session is completed (i.e. as soon as possible). The research does not involve any conflict of interests.

Briefing sheet (face construction)

Please keep this page for your information.

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites, and has received ethical approval.

If you agree to take part, your involvement will be in two stages. You will first be asked to look at a photograph of a UK football player, a face that will be unfamiliar to you. The following day, with my assistance, you will construct a composite of his face. This will take up to about an hour and a half. I am able to offer you £5 for your time for producing a composite. Alternatively, if you are a first- or second-year student, you can receive course credit.

Please be aware that you do not have to complete the experiment. This means that you can leave at any time from the start to the end of the experiment just by telling me that you wish to leave. This also means that you are free to leave between the session now and the session tomorrow (by phoning or sending me an email). At the end of the experiment, should you permit, your data will be combined with other people's data and analysed to see if the experimental aims have been met. (These aims will be explained to you at the end of tomorrow's session.) Your name and other identifying information will not be recorded, and so the data will remain anonymous. As it is anonymous, this means that after the experiment your data cannot be returned to you and will be used for analysis. The intention is to ultimately publish results in scientific journals and discuss them at academic conferences, to inform forensic theory and practice, but any data discussed will not be identifiable as yours (as no names or other identifying information were recorded).

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

Thank you,

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Briefing sheet (composite face naming)

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites, and has received ethical approval.

If you agree to take part, your involvement will be voluntary and will last for about 10 minutes. You will be presented with composites of well-known football players who compete at an international level in the UK. First, you will be asked to name as many of the identities as possible, or guess if unsure. I will also ask you to name the photographs used to construct the composites, as a check that you are familiar with each identity.

Other people will also be asked to name the composites. When this is done, the plan is to pool these data and analyse them using statistics to see if the experimental aims have been met (these will be explained to you at the end of today's session). In that analysis, your data will not be identifiable as it is anonymous; the analysis will also be based on group (not individual) data. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Any point during your involvement in the experiment, you have the right to withdraw and not complete the experiment. Just tell me this at any time; any data collected will then not be used. Please note that at the end of today's research session with you, it will not be possible to remove this data since it will not have your name associated with it and cannot be identified when mixed in with other people's data.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

Thank you,

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Briefing sheet (likeness rating)

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites, and has received ethical approval.

If you agree to take part, your involvement will be voluntary and will last for about 15 minutes. You will be presented with composites of football players who compete at an international level in the UK. I will ask you to give a rating of likeness for each composite image in the presence of the relevant target photograph. For each image, I'll ask you to give a rating from 1 (very-poor likeness) to 7 (very-good likeness).

Other people will also be asked to rate the composites in the same way. When this is done, the plan is to pool these data and analyse them using statistics to see if the experimental aims have been met (these will be explained to you at the end of today's session). In that analysis, your data will not be identifiable as it is anonymous; the analysis will also be based on group (not individual) data. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Any point during your involvement in the experiment, you have the right to withdraw and not complete the experiment. Just tell me this at any time; any data collected will then not be used. Please note that at the end of today's research session with you, it will not be possible to remove this data since it will not have your name associated with it and cannot be identified when mixed in with other people's data.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

Thank you,

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Debriefing sheet (face construction)

Please keep this page for your information.

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites and has received ethical approval.

In the study, you were in one of four conditions. If you were in the control condition, you would have constructed a composite in a similar way to eyewitnesses, by describing the target face using cognitive interviewing techniques and then guiding the artist to create a sketch of the target. In an alternative condition, you would have used the PRO-fit composite system to identify facial features for sketching. You also would have viewed the target for either 60 seconds or 5 seconds.

Your composite will be mixed in with composites from other participants in the study and then shown to other people who are familiar with the targets and asked to name them (and also give a rating of the likeness with the target face). These composite naming and rating data will be analysed to see if the aims of the experiment have been met.

As a result of this analysis, it is expected that composites will be more identifiable when constructors used PRO-fit composite system to select facial features in the context of a complete face after which a sketch was created. This is expected in both target exposure conditions but more of a benefit is expected for the 5 second exposure time. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Just to remind you that you have the right to have your composite withdrawn and thus not used for the following naming stage of the experiment. Please tell me now if you would like to do this. Otherwise, it will not be possible to remove the composite from the set since it will not have your name associated with it and cannot be identified when mixed in with other people's composites.

Thank you for participating. For further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Reference. General information about facial composites may be found in the following review chapter: Frowd (2014). Facial composite systems. In T. Valentine and J. Davis (Eds.) Forensic Facial Identification. Wiley-Blackwell. (It is available for download at http://www.evofit.co.uk/research/)

Debriefing sheet (face naming)

Please keep this page for your information.

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites and has received ethical approval.

In the study, you looked at sketch composites that were constructed by previous participants in one of four conditions. In the control condition, composites were constructed in a similar way to eyewitnesses, by describing the target face using cognitive interviewing techniques and then guiding the artist to create a sketch of the target. In an alternative condition, participants used the PRO-fit composite system to identify facial features for sketching. For all participants, the exposure to the target face was 5 or 60 seconds.

Your naming data will be mixed anonymously with data from other participants in the study. These composite naming data will then be analysed to see if the aims of the experiment have been met.

As a result of this analysis, it is expected that composites will be more identifiable when constructors used PRO-fit composite system to select facial features in the context of a complete face after which a sketch was created. This is expected in both target exposure conditions but more of a benefit is expected for the 5 second exposure time. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Just to remind you that you have the right to have your data withdrawn and thus not used for the following analysis stage of the experiment. Please tell me now if you would like to do this. Otherwise, your data will be mixed in anonymously with other people's data and so it will not be possible to identify it for retraction.

Thank you for participating. For further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Reference. General information about facial composites may be found in the following review chapter: Frowd (2014). Facial composite systems. In T. Valentine and J. Davis (Eds.) Forensic Facial Identification. Wiley-Blackwell. (It is available for download at http://www.evofit.co.uk/research/)

Debriefing sheet (likeness rating)

Please keep this page for your information.

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites and has received ethical approval.

In the study, you looked at sketch composites that were constructed by previous participants in one of four conditions. In the control condition, composites were constructed in a similar way to eyewitnesses, by describing the target face using cognitive interviewing techniques and then guiding the artist to create a sketch of the target. In an alternative condition, participants used the PRO-fit composite system to identify facial features for sketching. For all participants, the exposure to the target face was 5 or 60 seconds.

Your likeness-rating data will be mixed anonymously with data from other participants in the study. These data will then be analysed to see if the aims of the experiment have been met.

As a result of this analysis, it is expected that composites will be have higher likeness ratings when constructors used PRO-fit composite system to select facial features in the context of a complete face after which a sketch was created. This is expected in both target exposure conditions but more of a benefit is expected for the 5 second exposure time. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Just to remind you that you have the right to have your data withdrawn and thus not used for the following analysis stage of the experiment. Please tell me now if you would like to do this. Otherwise, your data will be mixed in anonymously with other people's data and so it will not be possible to identify it for retraction.

Thank you for participating. For further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Reference. General information about facial composites may be found in the following review chapter: Frowd (2014). Facial composite systems. In T. Valentine and J. Davis (Eds.) Forensic Facial Identification. Wiley-Blackwell. (It is available for download at http://www.evofit.co.uk/research/)

6.2 APPENDIX B – Unpublished paper

Improving the effectiveness of sketch-based composite images

Heidi Kuivaniemi-Smith (1*)

Charlie D. Frowd (1)

Hazel Sanderson (1)

Simra Minahil (1)

(1) Department of Psychology, University of Winchester, Winchester SO22 4NR, UK

* Corresponding author: Heidi Kuivaniemi-Smith, Department of Psychology, University of Winchester, Winchester SO22 4NR, UK. Email: Heidi.Kuivaniemi-Smith@winchester.ac.uk. Phone: 07979977334.

Running head: Refining sketch production

Introduction

Interview is an integral part of facial composite construction. The Association of Chief Police Officers guidelines in the UK and International Association of Identification (IAI) guidelines in the USA (ACPO, 2009; Richardson *et al*., 2010) recommend the Cognitive Interview to be used to facilitate (eye)witness recall in a context of composite construction. There are core mnemonics to be included in the interview but the interviewer needs to assess which others are appropriate to use in each case. See (Kuivaniemi-Smith et al., this volume) for further details on sketch artists' use of the cognitive interview and its application in research.

Describing a face is a very specific task and involves focusing on individual features. Featural composite systems used in the UK such as E-FIT and PRO-fit have been found to perform rather poorly in previous research in terms of producing identifiable composites (e.g. Frowd et al., 2000; Bruce et al., 2000; Frowd et al., 2005). This is partly due to faces being recognised better when the facial features are in a whole face context rather than as isolated features(e.g. Tanaka & Farah, 1993; Tanaka & Sengco, 1997). This body of research has led to the development of recognition focused holistic composite systems such as EvoFIT whose performance has continuously been improved and its identification rate has been demonstrated to be superior to that of the feature-based systems (see Frowd, 2014).

Focusing on the personality traits of a face instead of its physical features has been known to affect recognition positively (Berman and Cutler, 1989), and when this is done while encoding a face, it has also been found to influence composite quality (Shepherd, Ellis, McMurran and Davies, 1978; Wells and Hryciw, 1984; Davies and Oldman, 1999). This has led to the development of a holistic cognitive interview, which adds a part in the cognitive interview that aims to aid the holistic processing of a face. The witness is asked to think about the character of the offender's face for one minute and then to make character judgments (e.g. aggressiveness, intelligence) on a three-point scale (Frowd et al., 2008). This has been demonstrated to work efficiently with composite construction (Frowd et al., 2008, 2012, 2013); for example in Frowd et al. (2008), the correct naming of PRO-fit composites constructed after the CI alone was 9% and increased to 41% after the H-CI. The benefits were therefore evident even though the feature selection in PRO-fit was not holistic in nature but features were seen in isolation from each other.

H-CI was also superior to face recall CI using EvoFIT in Frowd, Nelson & Skelton *et al.* (2012), with correct naming of 39% in H-CI and 24% in CI. Feedback from the Police indicates that the H-CI works effectively for witnesses with good recall of an offender's face (Frowd et al. 2008). It should also work well with composite systems in regular use by UK police (E-FIT, EFIT-V, EvoFIT and PRO-fit) (Frowd et al. 2008). Field studies found that EvoFIT composites constructed in this way led to identification rate of 60% compared to 14% identification rate of E-FIT composites (Frowd et al. 2012).

The benefits of the H-CI have been limited to computerised composite systems (see Frowd, 2014) whereas pilot studies in using H-CI with sketching have failed to improve composite quality. Stops (2012) compared a composite sketching method using CI, H-CI and H-CI internal features first. No reference materials, in a form of a catalogue containing different facial features, were used in this study. No significant difference was found between the methods in a composite naming task using a Pearson's Chi-squared test. In CI the naming rate was 19.7% , in H-CI IF 18.9% and in H-CI 31%. It was unexpected that the H-CI IF condition did not prove to be superior to the other methods based on findings of previous research.

Although this type of sketching method (see more details on the sketching method – Fodarella et al., 2015)seems more holistic in nature than the featural systems since the initial sketch is shown to the witness only when the whole face has been drawn, the alteration of individual features from thereon can still be seen as focusing on individual features fairly heavily. Seeing different facial features in a context of a whole face is more closely-aligned to face recognition (Tanaka and Farah, 1993 and Davies and Christie, 1982) and perhaps using reference materials (pictures of facial features) in combination with sketching may benefit the face construction process. The modern feature systems such as PRO-fit and E-FIT are suitable for this as they allow features to be selected in a whole face context as well as individually.

A study by Skelton et al. (2015) employed PRO-fit's whole face option in composite construction. This method was compared to face construction selecting isolated features. The whole face condition was significantly better than the isolated. This finding was replicated in another naming task of this experiment which compared the quality of the internal features. This study offers further support for holistic recognition specifically in the context of computerised composite construction. It was concluded that composites constructed with featural systems should use the whole face option for feature selection instead of the isolated option.

The computerised systems and sketching have been compared to each other separately but not in combination with each other. Since the whole face feature selection has been found to be effective on PRO-fit, this and potential benefit of the flexibility of sketching should be explored.

The way the features are presented to the composite constructor was compared in Experiment 1, which combined PRO-fit and sketching. 2 interview methods (CI vs. H-CI) and 2 methods of feature selection (isolated vs. whole face) were tested. See Figure 1 for different feature selection methods. It was hypothesised that the whole face condition would produce more identifiable composites than the isolated feature condition and also that the composites in the H-CI condition would be superior to the composites in the CI condition.

Figure 1. Methods for selecting example pairs of eyes from PRO-fit: top row illustrates isolated-feature selection and bottom row illustrates feature selection in a whole-face context.

EXPERIMENT

Two stages were required to carry out the investigation: composite construction (Stage 1) and composite naming (Stage 2).

Stage 1: Composite Construction

Method

Design

In Stage 1, participants were shown a target individual, and were interviewed the following day by an experienced sketch artist (the researcher) to recall the given face using a CI or an H-CI were then advised to select facial features as isolated elements or in a whole-face context (see Figure 1). Having identified facial features, the artist worked with participants to draw a sketch of the face by hand. The design was 2 interview (CI vs. H-CI) x 2 method of feature selection (isolated vs. whole face) [1].

here. However, the technique (which was assessed in Stages 1 and 2 the same as the other conditions) turned out to be ineffective (as indicated by naming) and so, for the sake of brevity, we only mention it here.

Targets were photographs of footballers (see Materials). These images were shown to participants who were unfamiliar with the identities, to simulate the real-world situation where offenders are unknown to witnesses. In Stage 2, participants were recruited who were familiar with these identities (e.g. football fans), to be able to recognise (name) the composites constructed in Stage 1.

---

[1] We supplemented this between-subjects design with an experimental condition emerging from previous research. In that work, participants underwent an H-CI, and then constructed a composite using PRO-fit whereby they identified internal features (with external features masked) and then external features (with internal features masked). This part-face construction method was effective for constructing composites using this feature system and so we thought that it should allow accurate selection of facial features for sketching

Materials
Materials were 10 colour front-facing photographs of different male footballers who play at an international level in the UK. They were largely clean shaven and presented a neutral expression. Images were printed in colour to dimensions of 8cm (wide) x 10cm (height). PRO-fit software version 3.5.15 was used as the source of reference materials for selection of facial features.

Participants
Face constructors were volunteers living and working around the Thames Valley area. They were recruited on the basis of not following UK football. There were 17 males and 33 females with an age range from 17 to 70 (M = 42.32, SD = 14.7) years.

Procedure

Participants were tested individually by the researcher (first author). They were assigned to one of four conditions in a 2 interview (CI vs. H-CI) x 2 selection (isolated feature vs. whole-face) design; assignment of participants to condition and target was random within the constraint that the 10 targets were constructed once per condition. Participants were shown a target photograph from the assigned condition and were asked whether the face was recognised; if it was stated to be familiar, another face was selected randomly. For the first unfamiliar face reported, participants inspected the face for 60 seconds in the knowledge that a composite of it would be created the following day.

The experimenter and participant met 20 to 28 hours later to prepare a sketch. The procedure used by the researcher to recall the appearance of the face (for CI or H-CI) and then create the sketch was as described in Fodarella et al. (2015). In brief, participants were given an overview of the face-construction procedure; this briefing also included use of PRO-fit for selection of facial features following the assigned interview. After the relevant (CI / H-CI) interview had been administered, some more information about the features that had not been recalled in detail was probed, however kept minimal, so that the researcher was able to enter the given face description into PRO-fit, to allow approximately 20 features to be presented. Individual features were shown to constructors for identification of best likeness as per the assigned method of selection (isolated or whole-face). Constructors were encouraged to give additional detail to be taken into account when preparing the sketch. Once individual features had been identified, the sketch was constructed (as described by Fodarella et al.2015). Sketches took about one and a half hours to complete including the time for the interview, selection of features and debriefing.

Stage 2: Composite naming

Method

Design
Participants who claimed to be familiar with UK footballers were asked to name composites from one of the four conditions in Stage 1 and so the design was the same as above.

Materials
Materials were the 40 greyscale composites (10 from each method of construction) and the 10 target colour photographs. Images were printed to the same dimensions as for face construction.

CI	H-CI	CI	H-CI	
Isolated feature	Isolated feature	Whole face	Whole face	

Figure 2. Example composites constructed in the study of UK footballer, Frank Lampard. Each sketch was produced from memory by a different participant following one of two types of interview (CI or H-CI) and with facial features selected using one of two methods (isolated or whole face). For reasons of copyright, the actual target photograph cannot be reproduced here, but a sketch by the researcher has been included representing Frank Lampard.

Figure 3. A sketch of Frank Lampard.

Participants

Participants who named the composites were 25 male and five female volunteers from around the Thames Valley area. They were recruited on the basis of being familiar with footballers who play at an international level in the UK. Their age ranged from 22 to 70 (M = 42.5, SD = 13.2) years.

Procedure

Participants were tested individually. They were randomly assigned to one of the four sets of composites created in Stage 1. Participants were told that they would be presented with sketch composites of UK international-level footballers to name; it was mentioned that participants were free to make guesses or not give a name at all. The assigned set of sketches were presented sequentially and participants offered a name (or not). Next, participants were informed that viewing a composite from the side can improve face recognition and so were invited to have another go at naming, by inspecting each sketch in this way. Afterwards, the 10 target photographs were similarly presented sequentially and participants were asked to name these images. Each person received a different random order of presentation for composites and targets. The naming task took about 25 minutes to complete, including the time for debriefing.

Results

Participant responses (N = 600) to sketched composites were checked for missing data (of which no cases were found); they were scored for accuracy with respect to the relevant identity: a value of 1 was assigned when the correct name was given and 0 otherwise. Responses to target photographs (shown after participants have seen their assigned set of composites for the second time) were processed likewise. Most participants (*N* = 21 / 30) correctly named all 10 targets: in total, 288 responses were correct out of a possible 300 (*M* = 96.0%), indicating that participants were familiar with the relevant identities. However, failure to recognise a target suggests that participants could not have correctly identified the relevant composite, and so composites for these cases were coded as missing data. The resulting mean naming was 13.9% correct for composites. This level of performance is considerably less than naming of target photographs, but this is the usual situation as composites are error-prone stimuli that rarely (if ever) enjoy perfect recognition.

The composite naming data were subjected to Generalised Estimating Equations (GEE), a regression-type analysis that provides a combined by-participant and by-item model (Barnett et al., 2009). The three main predictors (type of interview, method of feature selection and view at naming) and their four interactions were subject to sequential backward elimination (based on $p > .1$ and lowest X^2). The within-subjects variable was view (at naming). The eight cells of the design were checked for appropriateness for this frequency type of analysis: *f(observed)* > 0, and *f(expected)* < 5 for at most 20% of cells. A robust estimator was used for the covariance matrix and an auto-regression structure for the working correlation matrix; the link function was binary logistic.

The final model [Intercept: *B* = -1.1, *SE(B)* = 0.2, $p < .001$, 1/*Exp(B)* = 2.9] emerged significant for method of feature selection [$X^2(1) = 5.0$, $p = .025$], with a large benefit for whole face (*M* = 20.5%) over isolated (*M* = 11.0%) [*B* = 1.7, *SE(B)* = 0.5, $p < .001$, *Exp(B)* = 5.8][(2)], and for the interaction between (i) interview and selection [$X^2(1) = 8.8$, $p = .003$] and (ii) interview and view [$X^2(2) = 11.2$, $p = .004$].

The two significant interactions were analysed using separate GEE, each containing one interaction to allow examination of the relevant pairwise parameter estimates (and without including other predictors). For interview x selection (Table 1), H-CI was superior to CI for whole-face selection [*B* = 0.6, *SE(B)* = 0.3, $p = .015$, *Exp(B)* = 1.9] but the opposite was found for isolated-feature selection, with CI emerging much better than H-CI [*B* = 1.4, *SE(B)* = 0.6, $p = .019$, *Exp(B)* = 4.2]. For interview x view (Table 2), the naming benefit of side over front view was restricted to sketches produced with the H-CI [*B* = 0.3, *SE(B)* = 0.2, $p = .027$, one tailed, *Exp(B)* = 1.3] (see also Table 2, Note).

Table 1. Correct naming of composites: interaction between interview and feature-selection method

	Interview	
Selection	CI	H-CI
Isolated feature	17.2[a] (20 / 116)	5.0 (6 / 120)
Whole face	14.3[a] (14 / 98)	25.0 (34 / 136)

[2] For readers unfamiliar with this type of analysis: (i) *B* is the coefficient of the predictor (slope of the regression line), (ii) *SE(B)* is the standard error of this coefficient, (iii) *Exp(B)* is the Odds Ratio, which is interpretable in terms of effect size with approximate values of 1.5 for small, 3.5 for medium and 9.0 for large. Note that, to aid interpretation, the opposite of the Odds Ratio is sometimes used, the Risk Ratio, to avoid presenting an effect size as a fractional quality.

Note. Values are percentage-correct accuracy calculated from responses in parentheses: summed correct responses (numerator) and total (correct and incorrect) responses (denominator). These data are for composites for which participants correctly named the relevant target (N = 588 out of 600). [a] Difference is not significant (p = .40), unlike all other pairwise comparisons (p < .02). See text for more details.

Table 2. Correct naming of composites: interaction between interview and view at naming

	Interview	
View at naming	CI	H-CI
Front	15.9[a] (17 / 107)	14.1† (18 / 128)
Side	15.9[a] (17 / 107)	17.2† (22 / 128)

Note. See Table 1 (Note) for definitions. †p < .05. [a] While these two means are identical, analysis of the interaction indicated that naming was better from front than side [B = 0.07, *SE(B)* = 0.03, p = .006]. This curious result is due to (i) view at naming being within-subjects, (ii) low *SE(B)* and (iii) the method of calculation: this was based on estimated marginal means, which was higher for front (M = 15%) than for side (M = 14%). Note that the effect size is very small [*Exp(B)* = 1.1] and so is of little practical importance anyway.

Composites are sometimes misidentified. When these errors occur, they may trigger false leads in a police investigation, and so analysing their pattern of occurrence is of practical value. In the current project incorrect names of composites were analysed as a proportion of the total responses from participants once correct responses had been removed (and also, as above, after scoring composites as missing data when targets were unknown). The number of mistaken names elicited was fairly frequent overall (N = 219 / 506), a mean of 43.3%. GEE were conducted in the same way as above on the incorrect naming responses. The model [Intercept: B = 0.1, *SE(B)* = 0.2, p = .62, *Exp(B)* = 1.1] revealed a single reliable estimator for view [$X^2(1) = 3.2$, p = .08], with slightly more incorrect names elicited when viewing the composite from the side (M = 44.6%) than the front (M = 42.0%) [B = 0.2, *SE(B)* = 0.1, *Exp(B)* = 1.2].

Experiment 2 — Replication of the H-CI advantage

Like experiment 1, this experiment included two stages: composite construction (Stage 1) and composite naming (Stage 2).

Stage 1: Composite Construction

Method

Design

In Stage 1, participants were shown a target individual, and were interviewed the following day by the same researcher than in experiment 1 to recall the given face using a CI and sketching without any reference materials or an H-CI and using PRO-fit composite system first to allow the participant to select facial features in a whole face context. The composite was then continued by sketching.

Targets were photographs of Coronation Street actors and actresses (see Materials). Like in Experiment 1, the composite constructors were unfamiliar with the targets and the participants attempting to name the composites were familiar with these identities.

Materials
Materials were 8 colour front-facing photographs of different actors and actresses of the Coronation Street TV- programme. Images were printed in colour to dimensions of 8cm (wide) x 10cm (height). PRO-fit software version 3.5.15 was used as the source of reference materials for selection of facial features.

Participants
Face constructors were volunteers living and working around the Thames Valley and Surrey area. They were recruited on the basis of not following Coronation Street. There were 6 males and 10 females with an age range from 21 to 67 (M = 45.3, SD = 15.2) years.

Procedure
The research design was the same as in experiment 1: 2 interview (CI vs. H-CI) x 2 face construction method (PRO-fit whole face vs. sketch); assignment of participants to condition and target was random within the constraint that the 8 targets were constructed once per condition. The participants' target view proceeded like in experiment 1 and the target viewing time was 60 seconds. The experimenter and participants met 20 to 28 hours later to prepare a sketch. The procedure used by the researcher to recall the appearance of the face (for CI or H-CI) and then create the sketch (PRO-fit whole face + sketch or sketch) was as described in Fodarella et al. (2015). Sketches took about one and a half hours to complete including the time for the interview and debriefing.

Stage 2: Composite naming

Method

Participants
Participants naming composites were sampled widely from town centres around Burnley, Nelson and Wakefield. There were 31 females and 13 males, and their age ranged from 21 to 79 (*M* = 39.0, *SD* = 16.3) years. Participants were volunteers and comprised an opportunity sample assigned equally to the two interview groups. A dozen participants failed the target-familiarity check (*a priori*) and were replaced to give the sample described here.

Materials, Design and Procedure
The basic procedure of Experiment 1 was used to name the composites except that we sought participants who were fans of Coronation Street, and told them that they would name composites of this type. Composites (no foils) were presented sequentially from one of two balanced sets (with random assignment to participants) comprising four composites from each of the two conditions in a within-subjects design; as before, the composite set was observed at a front view and then repeated at a side view. We applied an *a-priori* rule: participants must correctly name at least 75% of the target identities for their data to be included for analysis.

Results
The same basic procedure as Experiment 1 was followed for scoring, screening and analysing participant-naming responses. Correct naming of target photographs (carried out after naming of composites) was 94.4% overall and was thus similar to before (*M* = 96.0%). Accurate naming of composites was much higher in this study, at 43.8% correct (cf. *M* = 13.9%). For half of the targets, those that we constructed with low identification scores in the CI condition, correct naming of

composites markedly increased using the H-CI procedure, but declined for the other half. This observation indicates that ease of target-face construction (as measured by level of veridical [CI] naming) was related to effectiveness of the new interview procedure, as considered below.

These composite-naming data were subject to the same GEE with two within-subjects factors, type of interview (CI vs. H-CI) and view at naming (front vs. side). Interview was removed at Step 1 (p = .69), and the resulting model [Intercept: B = -2.1, $SE(B)$ = 0.2, p < .001, 1/$Exp(B)$ = 7.8] was significant for view [$X^2(1)$ = 39.0, p < .001], with a benefit of front over side [B = 0.6, $SE(B)$ = 0.1, p < .001, $Exp(B)$ = 1.9]. View interacted with interview [$X^2(2)$ = 6.4, p = .042] as the H-CI led to an increase in naming at the front view (cf. CI) but the opposite at the side view. We have observed interactions of this nature previously (e.g. Frowd et al., 2012; McIntyre et al., 2010) that have emerged due to level of vertical-composite naming. In this case, for composites created following a CI, mean naming was 38.3% for front but was much higher at 49.4% for side view, producing a crossover interaction. Consequently, we used coding level of mean-correct naming (M) in the veridical (CI) condition given that these data naturally separated into three mutually-exclusive equally-spaced groups: (Level 1) M < 35% (for 51% of the data), (Level 2) 35% < M < 65% (24% of the data) and (Level 3) M > 65% (25% of the data).

GEE were re-run using this (tri-)level coding for veridical naming. The model was reliable for interview x veridical naming [$X^2(5)$ = 98.9, p < .001]: There was a benefit of the H-CI procedure at Level 1 [B = 0.5, $SE(B)$ = 0.3, p = .048, one-tailed, $Exp(B)$ = 1.7], a null effect at Level 2 (p = .77) and a reliable deficit at Level 3 [-B = 1.2, $SE(B)$ = 0.4, p < .001, 1/$Exp(B)$ = 3.4]. Table 3 presents the marginal means by veridical naming and illustrates these reliable differences. It is worth mentioning that the main effect of interview was non-significant since the benefit of interview at Level 1 is half that of the deficit at Level 3, and that Level 3 contains about half the amount of data (providing a null effect overall); also, the dampening advantage of the H-CI is also evident in Experiment 1 with increasing naming in the CI, whole-face condition[3]. In practical terms, as discussed in the General Discussion, we argue that the lowest category in Experiment 2 is likely to be most prevalent in forensic practice, and that a reduction in performance at the top end is still very good anyway (and so likely to promote a recognisable image).

[3] In Experiment 1, mean naming was 0% correct for half of the target items in the CI (whole-face) condition, rising to 17.1% for these identities in the H-CI; for the other half, the increase in naming was considerably less, from 28.0 to 40.0%. (Note that accurate reliability statistics based on chi-square are intractable here due to lack of correct responses in the CI / whole-face condition.)

Table 3. Percentage-correct naming of composites: interaction between interview and level of veridical (CI) naming

	Interview	
Veridical (CI) naming	CI	H-CI
< 35	19.0† (3.7)	28.0† (3.9)
35 ... 65	50.0 (6.1)	48.0 (5.3)
> 65	88.0‡ (4.0)	69.0‡ (4.5)

Note. Values in parentheses are standard error. †p < .05. ‡p < .001.

Since veridical (CI) sketches varied widely by naming, from Level 1 to Level 3, we were curious as to whether a result reported by Frowd et al. (2013) would hold here. In that study, an advantage of view (i.e. an increase in correct naming from front to side) was much-more effective for composites that themselves were named well (M > 40%); it was argued that holistic (configural) information was more accurate in more-identifiable composites, facilitating recognition at the side view (which itself tends to hide featural errors but reveals configural aspects of the face). This observation should be apparent here: composites created using the H-CI at Level 3 (good-quality) should exhibit a stronger effect of view than at Level 1 (poor-quality).

So, GEE were run for the three-way interaction, which was reliable when included as a single term (p < .001). As predicted, the benefit of view was very effective at Level 3 [p < .001, Exp(B) = 6.3] but much weaker (and not significant) at Level 1 [p = .58, *Exp(B)* = 1.1] (since worse-quality composites are likely to have more-inaccurate configural relations). For composites created after a CI, the benefit of view was still apparent but much weaker [*Exp(B)* = 1.6] at Level 1 (p = .029); the effect size was about the same for good-quality composites [*Exp(B)* = 1.7] but not reliable (p < .18). These data for the CI condition suggest that sketch production itself is a somewhat holistic process, an observation that has been made previously (Davies and Little, 1990), since artists tend to work on groups of features when drawing the face.

Analysis of mistaken names were fairly infrequent overall (M = 12.3%). GEE revealed a significant interaction between interview and age [$X^2(1) = 6.1$, p = .013]: there was a decrease in mistaken names from CI to the new (H-CI) interview for a front view [B = -1.2, *SE(B)* = 0.5, p = .011, 1/*Exp(B)* = 3.4] but a roughly-equal increase from front to side view under H-CI [B = 1.8, *SE(B)* = 0.3, p = .001, *Exp(B)* = 3.2]. So, there were reliably more mistaken names given (from 3.6% to 10.9%) under the H-CI when naming from the side than the front, as found in Experiment 1.

Discussion

Summary.

The aim of experiment 1 was to explore how the combination of sketching and computerised composite system (PRO-fit) works in relation to facial features providing a recognition aid as reference images and being selected as isolated or in a whole face context. Another aim was to find out whether this new composite construction method would benefit from the holistic cognitive interview as previous studies have not found a benefit for the sketching method. The research design was 2 interview (CI vs. H-CI) x 2 method of feature selection (isolated vs. whole face)[4]. The face construction included two stages: viewing of the target face (unfamiliar to the participants) for 60 seconds and construction of the composite 20-28 hours after that. All the participants selected facial features in PRO-fit either as isolated or in a whole face context having described the face and been guided through the holistic interview in the H-CI condition. The artist then proceeded to create a sketch of the face by hand, working with the participant.

The aim of experiment 2 was to replicate the benefit of H-CI in the context of a sketching method. The most effective composite construction method of experiment 1 – H-CI PRO-fit whole face feature selection- was compared to the standard sketching method that uses no reference materials as recognition aids. The research design was similar to experiment 1: 2 interview (CI vs. H-CI) x 2 face construction method (PRO-fit whole face vs. sketch). The procedure was the same as in experiment 1, except this time the targets were Coronation Street actors and actresses. The standard sketching method proceeded as described in Fodarella et al. (this volume).

The hypotheses of the Experiment 1 were: 1) H-CI will lead to significantly more identifiable composites than CI in the whole face condition, 2) the whole face condition will lead to significantly more identifiable composites than the isolated feature condition in the CI conditions, 3) the side on view will enhance recognition of the composites significantly compared to the front view, 4) there will be no significant difference between the composites produced in the CI isolated feature condition and the H-CI isolated feature condition. The composites were evaluated by a recognition (naming) task by participants who were familiar with the target identities. Experiment 1 found that, as predicted, the whole face feature selection was significantly more effective than the isolated feature selection. There was also a significant interaction between the interview and feature selection with H-CI whole face condition emerging the most effective method and interview and view – the side on view eliciting significantly higher naming rate compared to the front view but this was restricted to the H-CI condition. The side on view has been claimed to correct the inaccuracies of the features and proportions by changing the aspect ratio of the face and therefore enhancing recognition (Frowd et al., 2013). This did not occur in the CI condition. Since the H-CI has been found to enhance holistic recognition of a face (Frowd et al., 2008), the quality of the composites is already likely to be better in the whole face condition than in the CI conditions making the side on view more effective. H-CI was superior to CI for whole-face selection but the opposite was found for isolated-feature selection, with CI emerging much better than H-CI. This finding could be explained by transfer- appropriate processing (Schooler et al., 1997). The CI tends to focus on featural information more, hence it works for isolated feature selection, and the H-CI enhances holistic recognition, therefore working better for whole face selection and clashing with isolated feature selection. This could potentially confuse the facial recognition process leading to low quality composites.

In experiment 2, it was hypothesised that the H-CI PRO-fit condition would produce significantly more identifiable composites than the other conditions. This was not the case in the overall naming

---

[4] We supplemented this between-subjects design with an experimental condition emerging from previous research. In that work, participants underwent an H-CI, and then constructed a composite using PRO-fit whereby they identified internal features (with external features masked) and then external features (with internal features masked). This part-face construction method was effective for constructing composites using this feature system and so we thought that it should allow accurate selection of facial features for sketching here. However, the technique (which was assessed in Stages 1 and 2 the same as the other conditions) turned out to be ineffective (as indicated by naming) and so, for the sake of brevity, we only mention it here.

rates. However, as the naming data was split into three sections- low, medium and high, half of the data was in the lower naming section.

TO BE CONTINUED…

To reference this article: Kuivaniemi-Smith, H., Sanderson, H., Minahil, S. and Frowd, C.D., “Improving the effectiveness of sketch-based composite images”, unpublished manuscript.

6.3 APPENDIX C – Examples of participant information sheets and a consent form (Experiment 2)

Briefing sheet (composite face naming)

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites, and has received ethical approval.

If you agree to take part, your involvement will be voluntary and will last for about 10 minutes. You will be presented with composites of well-known football players who compete at an international level in the UK. First, you will be asked to name as many of the identities as possible, or guess if unsure. I will also ask you to name the photographs used to construct the composites, as a check that you are familiar with each identity.

Other people will also be asked to name the composites. When this is done, the plan is to pool these data and analyse them using statistics to see if the experimental aims have been met (these will be explained to you at the end of today's session). In that analysis, your data will not be identifiable as it is anonymous; the analysis will also be based on group (not individual) data. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Any point during your involvement in the experiment, you have the right to withdraw and not complete the experiment. Just tell me this at any time; any data collected will then not be used. Please note that at the end of today's research session with you, it will not be possible to remove this data since it will not have your name associated with it and cannot be identified when mixed in with other people's data.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

Thank you,

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Briefing sheet (likeness rating)

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites, and has received ethical approval.

If you agree to take part, your involvement will be voluntary and will last for about 15 minutes. You will be presented with composites of football players who compete at an international level in the UK. I will ask you to give a rating of likeness for each composite image in the presence of the relevant target photograph. For each image, I'll ask you to give a rating from 1 (very-poor likeness) to 7 (very-good likeness).

Other people will also be asked to rate the composites in the same way. When this is done, the plan is to pool these data and analyse them using statistics to see if the experimental aims have been met (these will be explained to you at the end of today's session). In that analysis, your data will not be identifiable as it is anonymous; the analysis will also be based on group (not individual) data. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Any point during your involvement in the experiment, you have the right to withdraw and not complete the experiment. Just tell me this at any time; any data collected will then not be used. Please note that at the end of today's research session with you, it will not be possible to remove this data since it will not have your name associated with it and cannot be identified when mixed in with other people's data.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

Thank you,

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Debriefing sheet (face naming)

Please keep this page for your information.

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites and has received ethical approval.

In the study, you looked at sketch composites that were constructed by previous participants in one of four conditions. In the control condition, composites were constructed in a similar way to eyewitnesses, by describing the target face using cognitive interviewing techniques and then guiding the artist to create a sketch of the target. In an alternative condition, participants used the PRO-fit composite system to identify facial features for sketching. For all participants, the exposure to the target face was 5 or 60 seconds.

Your naming data will be mixed anonymously with data from other participants in the study. These composite naming data will then be analysed to see if the aims of the experiment have been met.

As a result of this analysis, it is expected that composites will be more identifiable when constructors used PRO-fit composite system to select facial features in the context of a complete face after which a sketch was created. This is expected in both target exposure conditions but more of a benefit is expected for the 5 second exposure time. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Just to remind you that you have the right to have your data withdrawn and thus not used for the following analysis stage of the experiment. Please tell me now if you would like to do this. Otherwise, your data will be mixed in anonymously with other people's data and so it will not be possible to identify it for retraction.

Thank you for participating. For further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Reference. General information about facial composites may be found in the following review chapter: Frowd (2014). Facial composite systems*. In T. Valentine and J. Davis (Eds.) Forensic Facial Identification*. Wiley-Blackwell. (It is available for download at http://www.evofit.co.uk/research/)

Frowd, C. D., Carson, D., Ness, H., McQuiston-Surrett, D., Richardson, J., Baldwin, H. and Hancock, P. (2005), Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology*, 10, 63–81. Retrieved from: 10.1348/135532504X15358

Questions:

1) As all the facial composites are sketches, do you think they have to be shown to participants as separate conditions or can the composites from different conditions be mixed? (see Frowd at al. 2005 for further info)
2) Why is a composite naming task considered the best way of evaluating the quality of the composites?

Debriefing sheet (likeness rating)

Please keep this page for your information.

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites and has received ethical approval.

In the study, you looked at sketch composites that were constructed by previous participants in one of four conditions. In the control condition, composites were constructed in a similar way to eyewitnesses, by describing the target face using cognitive interviewing techniques and then guiding the artist to create a sketch of the target. In an alternative condition, participants used the PRO-fit composite system to identify facial features for sketching. For all participants, the exposure to the target face was 5 or 60 seconds.

Your likeness-rating data will be mixed anonymously with data from other participants in the study. These data will then be analysed to see if the aims of the experiment have been met.

As a result of this analysis, it is expected that composites will be have higher likeness ratings when constructors used PRO-fit composite system to select facial features in the context of a complete face after which a sketch was created. This is expected in both target exposure conditions but more of a benefit is expected for the 5 second exposure time. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Just to remind you that you have the right to have your data withdrawn and thus not used for the following analysis stage of the experiment. Please tell me now if you would like to do this. Otherwise, your data will be mixed in anonymously with other people's data and so it will not be possible to identify it for retraction.

Thank you for participating. For further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries and concerns, please feel free to contact me or my supervisor using the following details.

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Reference. General information about facial composites may be found in the following review chapter: Frowd (2014). Facial composite systems. In T. Valentine and J. Davis (Eds.) Forensic Facial

Identification. Wiley-Blackwell. (It is available for download at http at
http://www.evofit.co.uk/research/)

Psychology Department

CONSENT FORM (**PARTICIPANT** COPY)

Investigating the effects of target exposure time and using reference materials (pictures of facial features) in facial composite construction.

I have read *(or had clearly explained to me)* and understood the information about the project. I understand that my participation in this project, is completely voluntary, and that I may withdraw at any time during the project, without penalty. I understand the arrangements that have been made to ensure my anonymity and privacy and that my data will remain anonymised.

The researcher has made clear to me any risks which may be involved in my participation in the project. On this basis, I consent to take part in the project.

Signed Participant Date

Signed Researcher Date

- -

Psychology Department

CONSENT FORM (**RESEARCHER** COPY)

Investigating the effects of target exposure time and using reference materials (pictures of facial features) in facial composite construction.

I have read *(or had clearly explained to me)* and understood the information about the project. I understand that my participation in this project, is completely voluntary, and that I may withdraw at any time during the project, without penalty. I understand the arrangements that have been made to ensure my anonymity and privacy and that my data will remain anonymised.

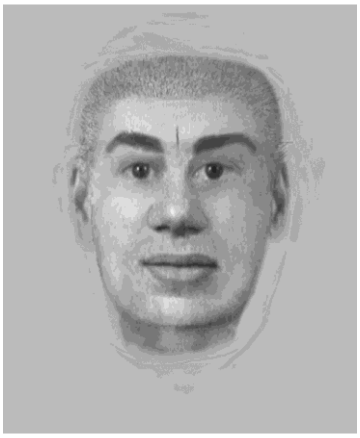
The researcher has made clear to me any risks which may be involved in my participation in the project. On this basis, I consent to take part in the project.

Signed Participant Date

Signed Researcher Date

6.4 APPENDIX D – composites from Experiment 1

Target: Jeroen van Koningsbrugge



Interview 30 (5 s PRO-fit)

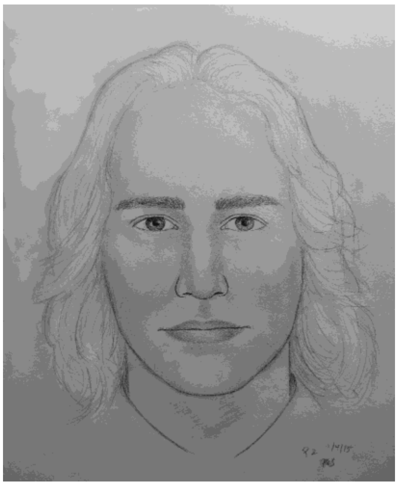
Interview 36 (5 s sketch)

Interview 1 (30 s PRO-fit)

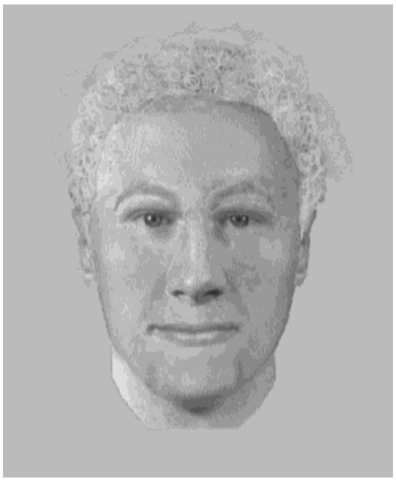
Interview 4 (30 s sketch)

Target: Gordon Heuckeroth

Interview 25 (5 s PRO-fit)



Interview 2 (5 s sketch)



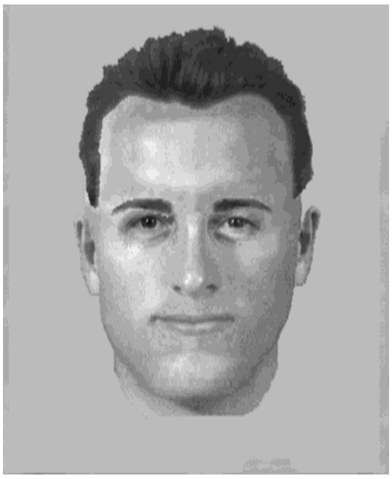
Interview 37 (30 s PRO-fit)

Interview 12 (30 s sketch)

Target: Frans Bauer

Interview 3 (5 s PRO-fit)

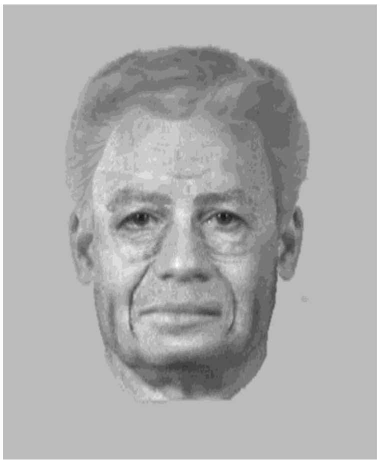
Interview 31 (5 s sketch)



Interview 29 (30 s PRO-fit)

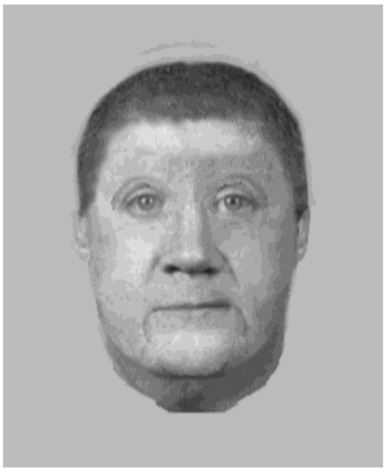
Interview 33 (30 s sketch)

Target: Bram Moskowicz



Interview 5 (5 s PRO-fit)

Interview 9 (5 s sketch)



Interview 11 (30 s PRO-fit)

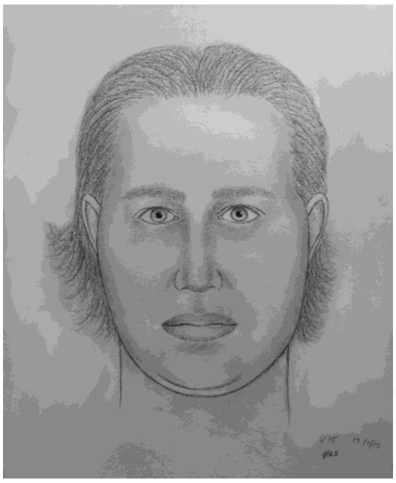
Interview 7 (30 s sketch)

Target: Marco Borsato

Interview 21 (5 s PRO-fit)

Interview 27 (5 s sketch)

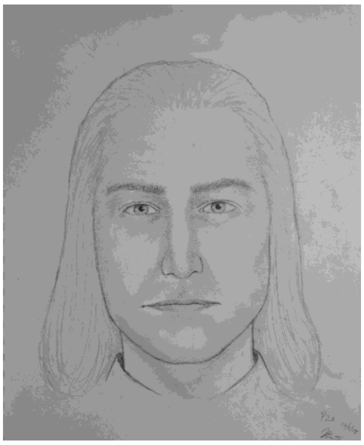
Interview 6 (30 s PRO-fit)



Interview 15 (30 s sketch)

Target: Matthijs van Nieuwkerk

Interview 8 (5 s PRO-fit)

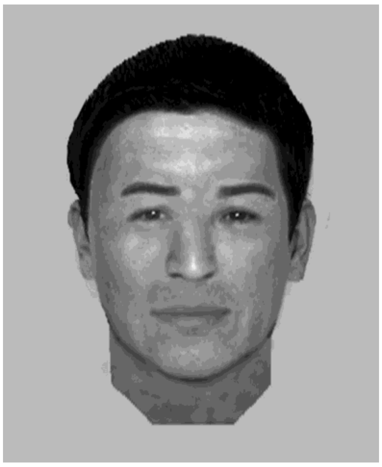


Interview 20 (5 s sketch)

Interview 26 (30 s PRO-fit)

Interview 39 (30 s sketch)

Target: Jan Smit



Interview 13 (5 s PRO-fit)

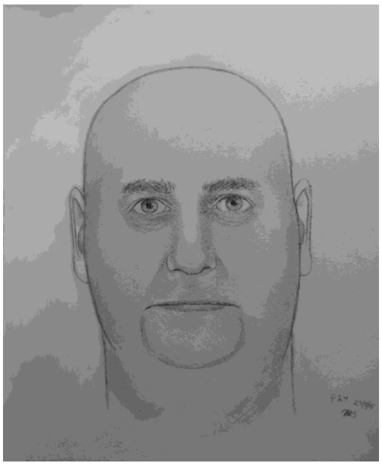
Interview 10 (5 s sketch)

Interview 18 (30 s PRO-fit)

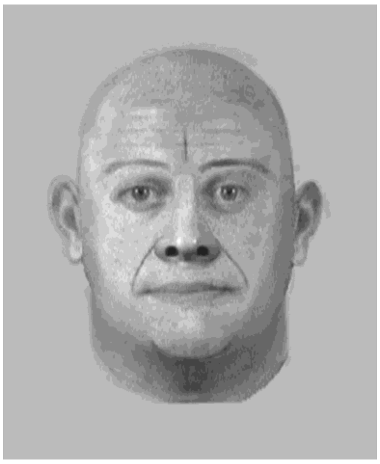
Interview 28 (30 s sketch)

Target: Paul de Leeuw

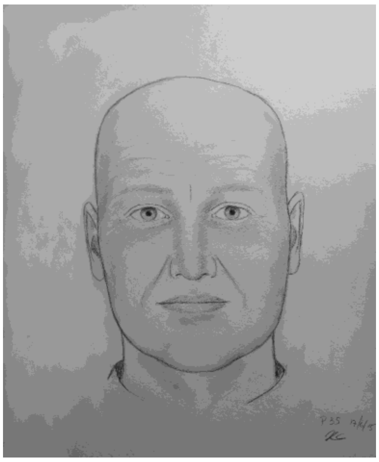
Interview 34 (5 s PRO-fit)



Interview 24 (5 s sketch)



Interview 14 (30 s PRO-fit)

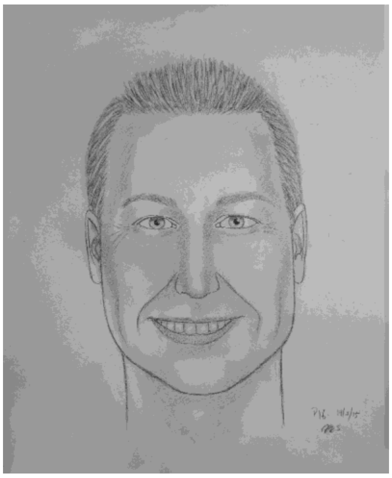


Interview 35 (30 s sketch)

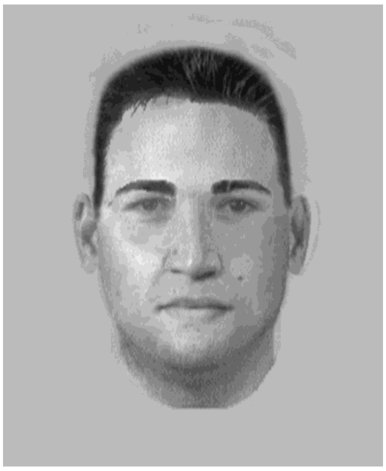
I

Target: Nick Schilder

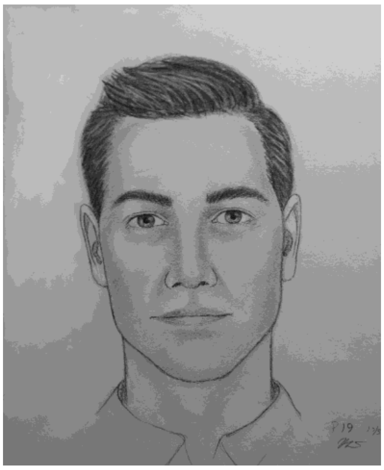
Interview 40 (5 s PRO-fit)



Interview 16 (5 s sketch)

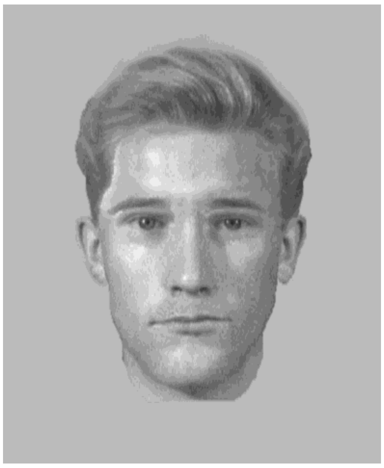


Interview 22 (30 s PRO-fit)



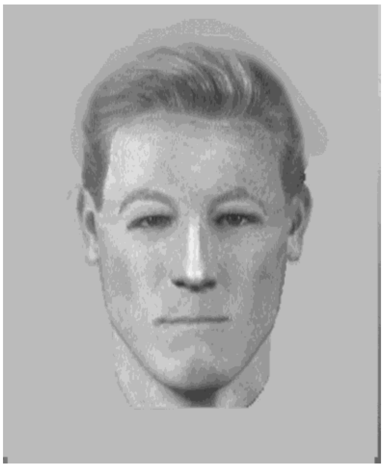
Interview 19 (30 s sketch)

Target: Dennis Storm



Interview 17 (5 s PRO-fit)

Interview 38 (5 s sketch)



Interview 32 (30 s PRO-fit)

Interview 23 (30 s sketch)

APPENDIX E – composites from Experiment 2

Target: Gareth Barry

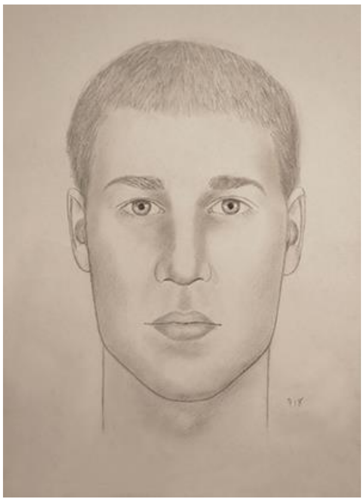
A interview 30 (5 s sketch CI)

C interview 39 (5 s PRO-fit/sketch CI)

B interview 9 (60 s sketch CI)

D interview 28 (60 s PRO-fit/sketch CI)

Target: Michael Carrick



A interview 18 (5 s sketch CI)

C interview 34 (5 s PRO-fit/sketch CI)

B interview 22 (60 s sketch CI)

D interview 10 (60 s PRO-fit/sketch CI)

Target: Peter Crouch

A interview 20 (5 s sketch CI)

C interview 3 (5 s PRO-fit/sketch CI)

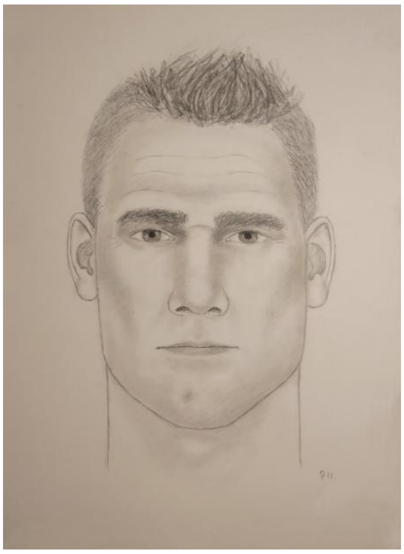
B interview 35 (60 s sketch CI)

D interview 6 (60 s PRO-fit/sketch CI)

Target: Steven Gerrard

A interview 1 (5 s sketch CI)

C interview 8 (5 s PRO-fit/sketch CI)



B interview 11 (60 s sketch CI)

D interview 23 (60 s PRO-fit/sketch CI)

Target: Frank Lampard

A interview 40 (5 s sketch CI)

C interview 24 (5 s PRO-fit/sketch CI)

B interview 31 (60 s sketch CI)

D interview 25 (60 s PRO-fit/sketch CI)

Target: James Milner

A interview 32 (5 s sketch CI)

C interview 21 (5 s PRO-fit/sketch CI)

B interview 27 (60 s sketch CI)

D interview 17 (60 s PRO-fit/sketch CI)

Target: Scott Parker

A interview 7 (5 s sketch CI)

C interview 37 (5 s PRO-fit/sketch CI)

B interview 15 (60 s sketch CI)

D interview 16 (60 s PRO-fit/sketch CI)

Target: Andy Carroll

A interview 26 (5 s sketch CI)

C interview 29 (5 s PRO-fit/sketch CI)

B interview 4 (60 s sketch CI)

D interview 14 (60 s PRO-fit/sketch CI)

Target: Ryan Giggs

A interview 19 (5 s sketch CI)

C interview 33 (5 s PRO-fit/sketch CI)

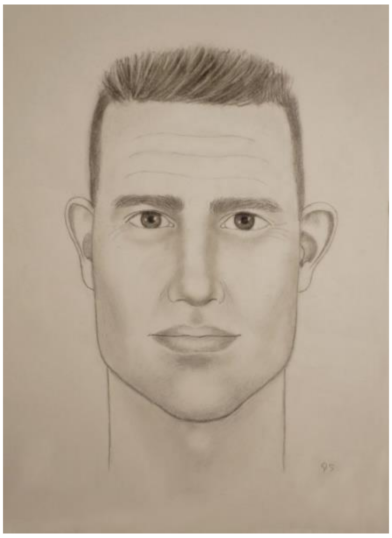
B interview 13 (60 s sketch CI)

D interview 38 (60 s PRO-fit/sketch CI)

Target: Robin Van Persie

A interview 2 (5 s sketch CI)

C interview 12 (5 s PRO-fit/sketch CI)



B interview 5 (60 s sketch CI)

D interview 36 (60 s PRO-fit/sketch CI)

6.5 APPENDIX F – composites from Experiment 3

Target: Chris Gunter

C Interview 29 (5 sec Ref)

A Interview 15 (5 sec No Ref)

D Interview 38 (60 sec Ref)

B Interview 31 (60 sec No Ref)

Target: Yann Kermorgant

C Interview 37 (5 sec Ref)

A Interview 25 (5 sec No Ref)

D Interview 3 (60 sec Ref)

B Interview 24 (60 sec No Ref)

Target: Roy Beerens

C Interview 12 (5 sec Ref)

A Interview 18 (5 sec No Ref)

D Interview 1 (60 sec Ref)

B Interview 27 (60 sec No Ref)

Target: George Evans

C Interview 34 (5 sec Ref)

A Interview 2 (5 sec No Ref)

D Interview 5 (60 sec Ref)

B Interview 6 (60 sec No Ref)

Target: John Swift

C Interview 30 (5 sec Ref)

A Interview 22 (5 sec No Ref)

D Interview 26 (60 sec Ref)

B Interview 17 (60 sec No Ref)

Target: Paul McShane

C Interview 10 (5 sec Ref)

A Interview 4 (5 sec No Ref)

D Interview 11 (60 sec Ref)

B Interview 36 (60 sec No Ref)

Target: Joey van den Berg

C Interview 13 (5 sec Ref)

A Interview 19 (5 sec No Ref)

D Interview 7 (60 sec Ref)

B Interview 14 (60 sec No Ref)

Target: Liam Kelly

C Interview 20 (5 sec Ref)

A Interview 9 (5 sec No Ref)

D Interview 32 (60 sec Ref)

B Interview 8 (60 sec No Ref)

Target: Vito Mannone

C Interview 28 (5 sec Ref)

A Interview 39 (5 sec No Ref)

D Interview 21 (60 sec Ref)

B Interview 33 (60 sec No Ref)

Target: David Edwards

C Interview 16 (5 sec Ref)

A Interview 40 (5 sec No Ref)

D Interview 35 (60 sec Ref)

B Interview 23 (60 sec No Ref)

6.6 APPENDIX G – composites from Experiment 4

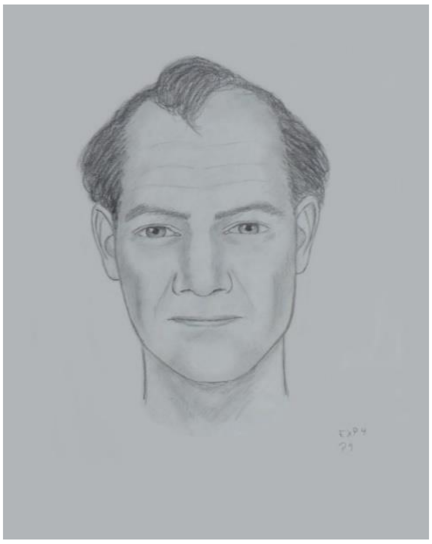
Target: Ian Beale

Interview 22 Min MRC

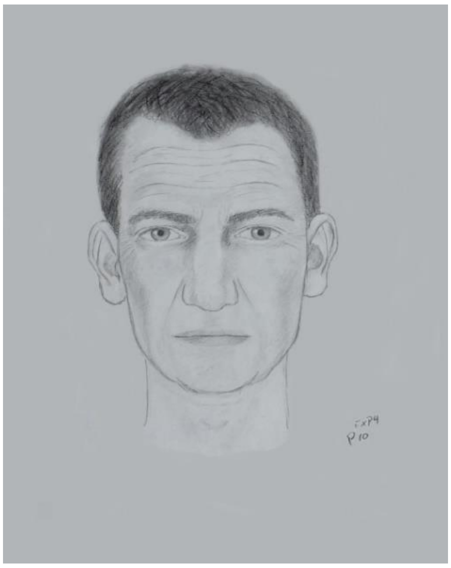
Interview 12 PRC

Interview 29 Detailed MRC

Target: Billy Mitchell



Interview 9 Min MRC



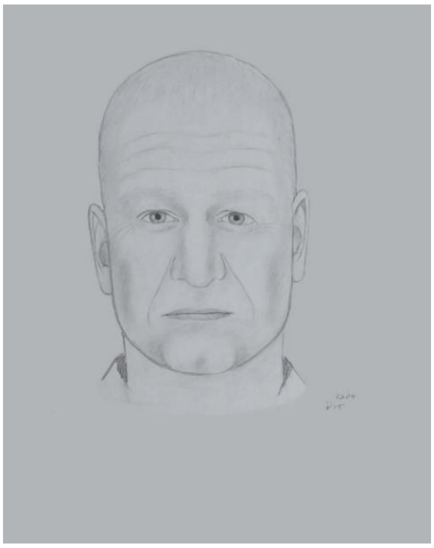
Interview 10 PRC

Interview 11 Detailed MRC

Target: Max Branning

Interview 23 Min MRC

Interview 18 PRC

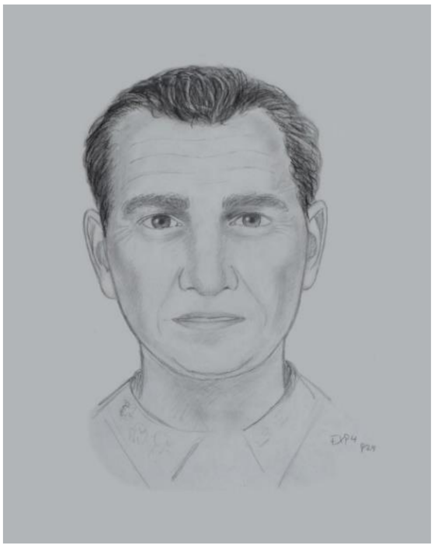


Interview 15 Detailed MRC

Target: Alfie Moon

Interview 21 Min MRC

Interview 7 PRC



Interview 24 Detailed MRC

Target: Mick Carter

Interview 4 Min MRC

Interview 8 PRC

Interview 13 Detailed MRC

Target: Stacey Slater/Branning

Interview 2 Min MRC

Interview 27 PRC

Interview 14 Detailed MRC

Target: Ronnie Mitchell

Interview 16 Min MRC

Interview 20 PRC

Interview 30 Detailed MRC

Target: Lauren Branning

Interview 17 Min MRC

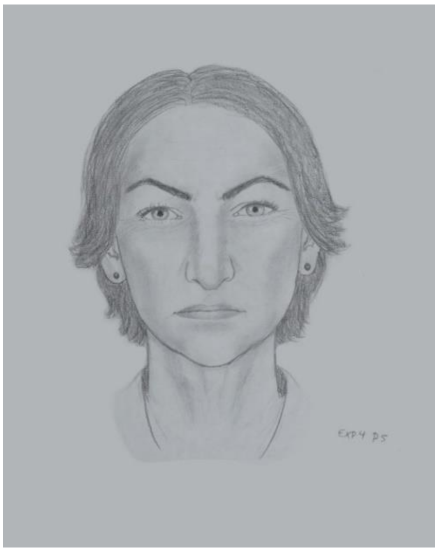
Interview 19 PRC

Interview 26 Detailed MRC

Target: Kat Slater

Interview 6 Min MRC

Interview 3 PRC



Interview 5 Detailed MRC

Target: Carol Jackson

Interview 28 Min MRC

Interview 25 PRC

Interview 1 Detailed MRC

6.7 APPENDIX H – composites from Experiment 5

Target 1 Debbie Dingle

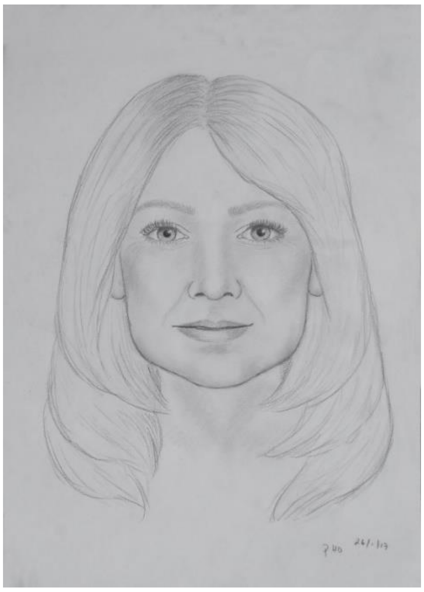
A Interview 11, min MRC sketch

B Interview 4, Detailed MRC sketch

C Interview 27, min MRC PRO-fit

D Interview 6, Detailed MRC PRO-fit

Target 2 Charity Sharma



A Interview 40, min MRC sketch

B Interview 17, Detailed MRC sketch

C Interview 3, min MRC PRO-fit

D Interview 21, Detailed MRC PRO-fit

Target 3 Chas Dingle

A Interview 15, min MRC sketch

B Interview 36, Detailed MRC sketch

C Interview 20, min MRC PRO-fit

D Interview 33, Detailed MRC PRO-fit

Target 4 Bernice Blackstock

A Interview 31, min MRC sketch

B Interview 23, Detailed MRC sketch

C Interview 34, min MRC PRO-fit

D Interview 29, Detailed MRC PRO-fit

Target 5 Moira Dingle

A Interview 8, min MRC sketch

B Interview 14, Detailed MRC sketch

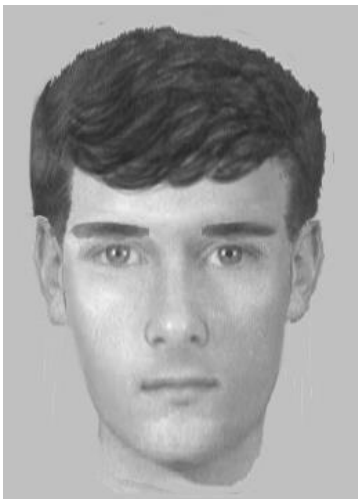
C Interview 30, min MRC PRO-fit

D Interview 13, Detailed MRC PRO-fit

Target 6 Robert Sugden

A Interview 7, min MRC sketch

B Interview 5, Detailed MRC sketch



C Interview 10, min MRC PRO-fit

D Interview 18, Detailed MRC PRO-fit

Target 7 Aaron Livesy

A Interview 16, min MRC sketch

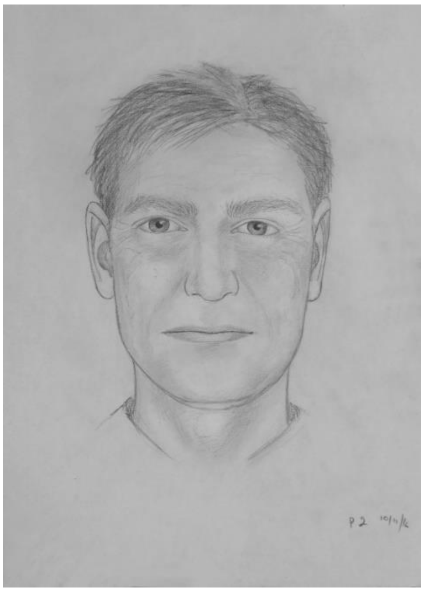
B Interview 12, Detailed MRC sketch

C Interview 25, min MRC PRO-fit

D Interview 26, Detailed MRC PRO-fit

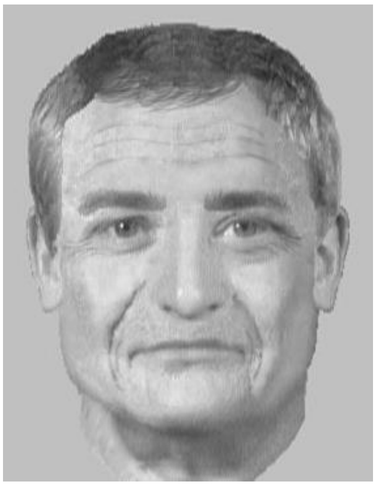
Target 8 Ashley Thomas

A Interview 19, min MRC sketch



B Interview 2, Detailed MRC sketch

C Interview 35, min MRC PRO-fit



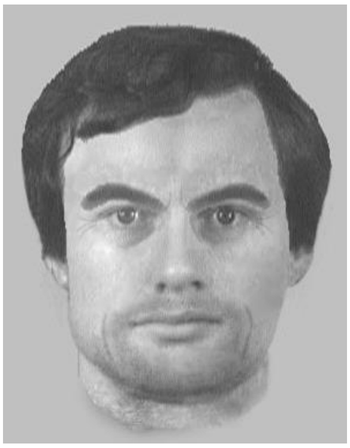
D Interview 9, Detailed MRC PRO-fit

Target 9 Cain Dingle

A Interview 1, min MRC sketch

B Interview 22, Detailed MRC sketch

C Interview 24, min MRC PRO-fit



D Interview 32, Detailed MRC PRO-fit

Target 10 David Metcalfe

A Interview 28, min MRC sketch

B Interview 37, Detailed MRC sketch

C Interview 39, min MRC PRO-fit

D Interview 38, Detailed MRC PRO-fit

6.8 APPENDIX I – Examples of the likeness rating task carried out in person with participants and online.

Experiment 2 – collecting data for likeness ratings in person with participants

Experiment 2 likeness rating task was completed face to face with participants, who were briefed about the study both verbally and in a written format as example above shows. After participants had signed the consent form, the first set of four composites and the corresponding target were taken out from an envelope and laid out on the desk in front of them (see Fig 1). The images were cut to same sized cards, and they were always shuffled after a participant had completed evaluating the composites and put back into the envelope. The order of the targets was randomised using an excel sheet (see Table 1). This procedure continued for all ten targets and their four composites. Each participant thus rated the likeness of 40 composites. They were instructed to tell the researcher if they wished to have a break during the task. The participants were told that the composites had been sketched from other participants' memory and descriptions and therefore were not portraits. They were instructed to base their ratings on the overall likeness and not analyse each individual feature for long. The task did not usually take longer than 15 minutes, however, some took longer than that and some were very fast.

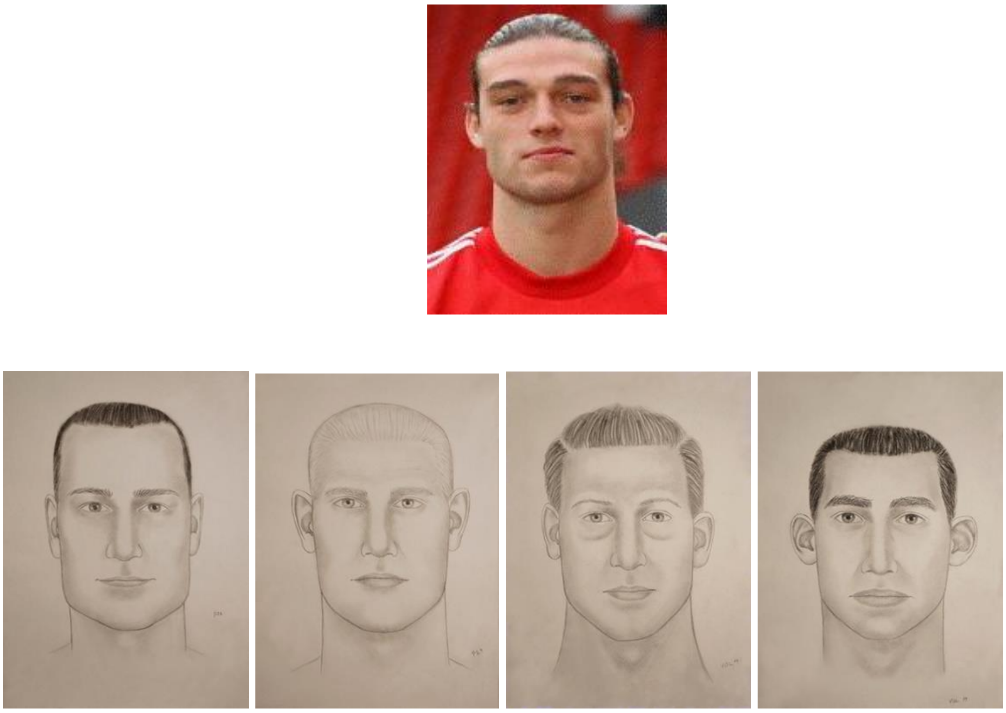


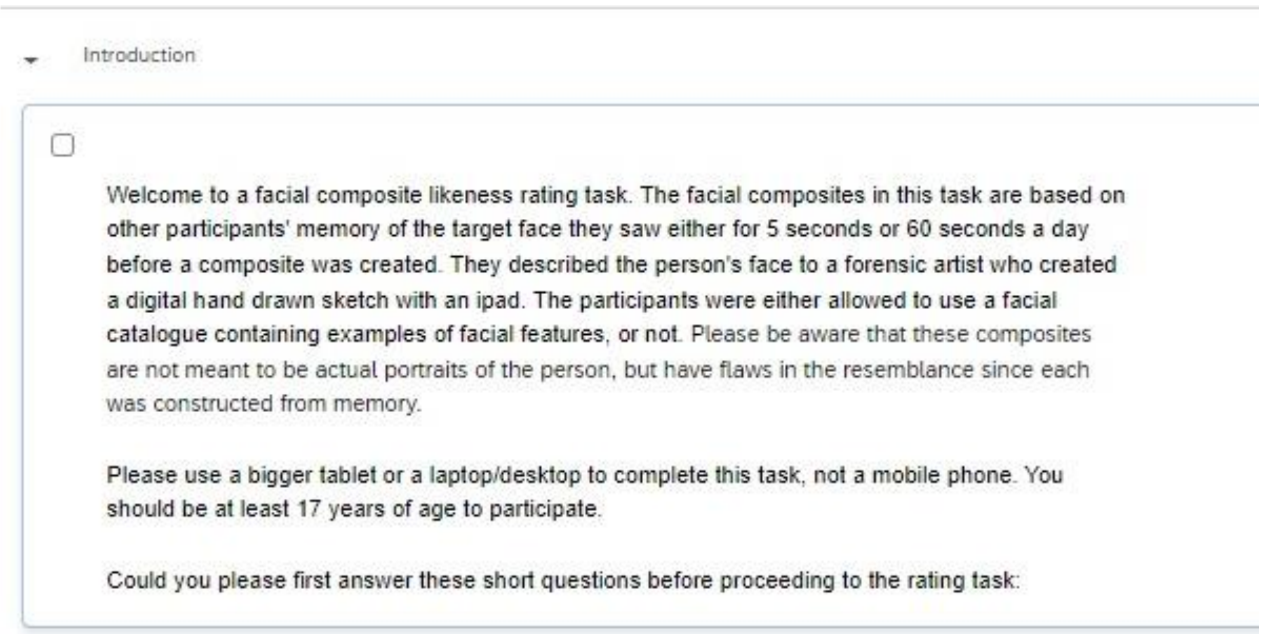
Fig 1. An example of how the composites and their corresponding targets were presented to the participants in face-to-face data collection.

Participant Target	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	6	5	1	9	10	2	4	3	9	1	10	7	8	6	7	8	5	4	3
2	10	10	6	2	1	2	8	8	5	8	2	4	5	4	5	1	6	4	6	5
3	8	7	2	5	7	5	6	10	6	2	4	5	9	2	3	8	7	9	1	2
4	4	1	1	4	10	3	3	9	2	7	5	6	3	7	9	10	2	3	7	10
5	3	2	9	6	4	9	5	5	7	10	3	7	1	6	7	3	4	6	2	4
6	6	5	4	3	3	4	9	2	8	5	8	1	10	3	4	5	9	2	9	8
7	7	9	3	7	2	1	10	3	10	6	9	3	6	5	1	4	1	8	5	6
8	1	4	8	10	5	8	4	6	1	1	7	2	8	9	10	9	10	7	8	7
9	9	8	10	9	6	7	1	7	4	3	10	8	2	1	8	2	3	1	10	9
10	5	3	7	8	8	6	7	1	9	4	6	9	4	10	2	6	5	10	3	1

Table 1. Target order randomised before presenting the target-composite sets to participants.

Experiment 3 – collecting data for likeness ratings online on Qualtrics survey site.

The likeness rating task was set up as an online survey on Qualtrics site. A brief explanation of the study was provided (see Fig 2), and the volunteers were required to give their consent to take part in the study if they wished to proceed to rating the composites (see Fig 3, public version of the survey). The survey questions were required to be answered before the next question was activated to avoid incomplete data. Randomisation was applied to the targets and their composites, so that they appeared in different order each time (see Fig 4).



Introduction

Welcome to a facial composite likeness rating task. The facial composites in this task are based on other participants' memory of the target face they saw either for 5 seconds or 60 seconds a day before a composite was created. They described the person's face to a forensic artist who created a digital hand drawn sketch with an ipad. The participants were either allowed to use a facial catalogue containing examples of facial features, or not. Please be aware that these composites are not meant to be actual portraits of the person, but have flaws in the resemblance since each was constructed from memory.

Please use a bigger tablet or a laptop/desktop to complete this task, not a mobile phone. You should be at least 17 years of age to participate.

Could you please first answer these short questions before proceeding to the rating task:

Fig 2. Study explained briefly at the beginning of the survey.

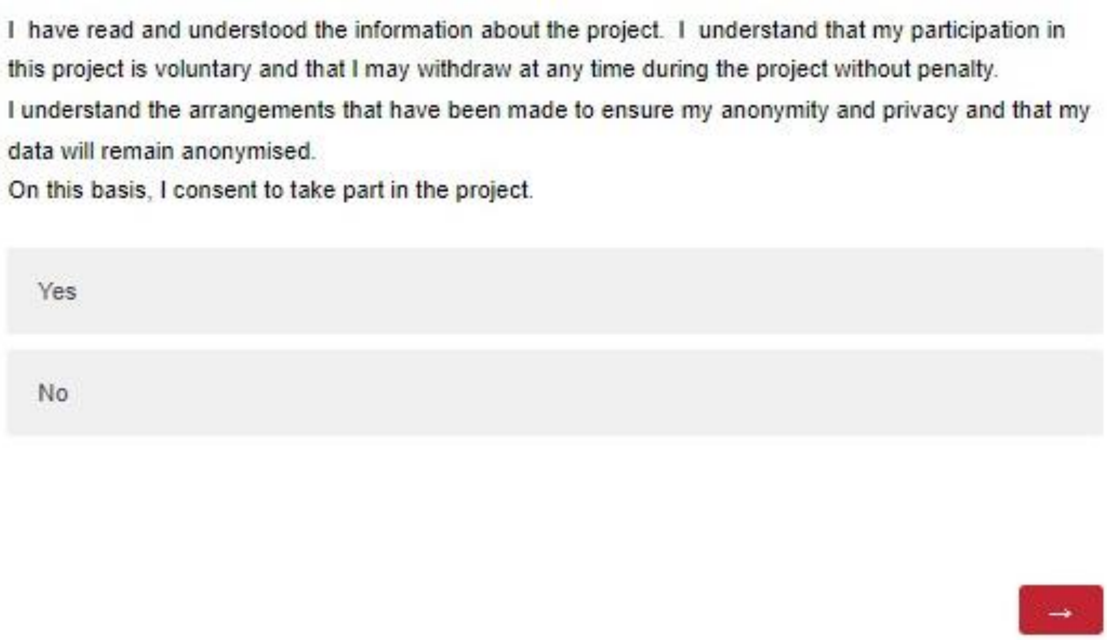


Fig 3. Consent from the participants was required before proceeding with the study.

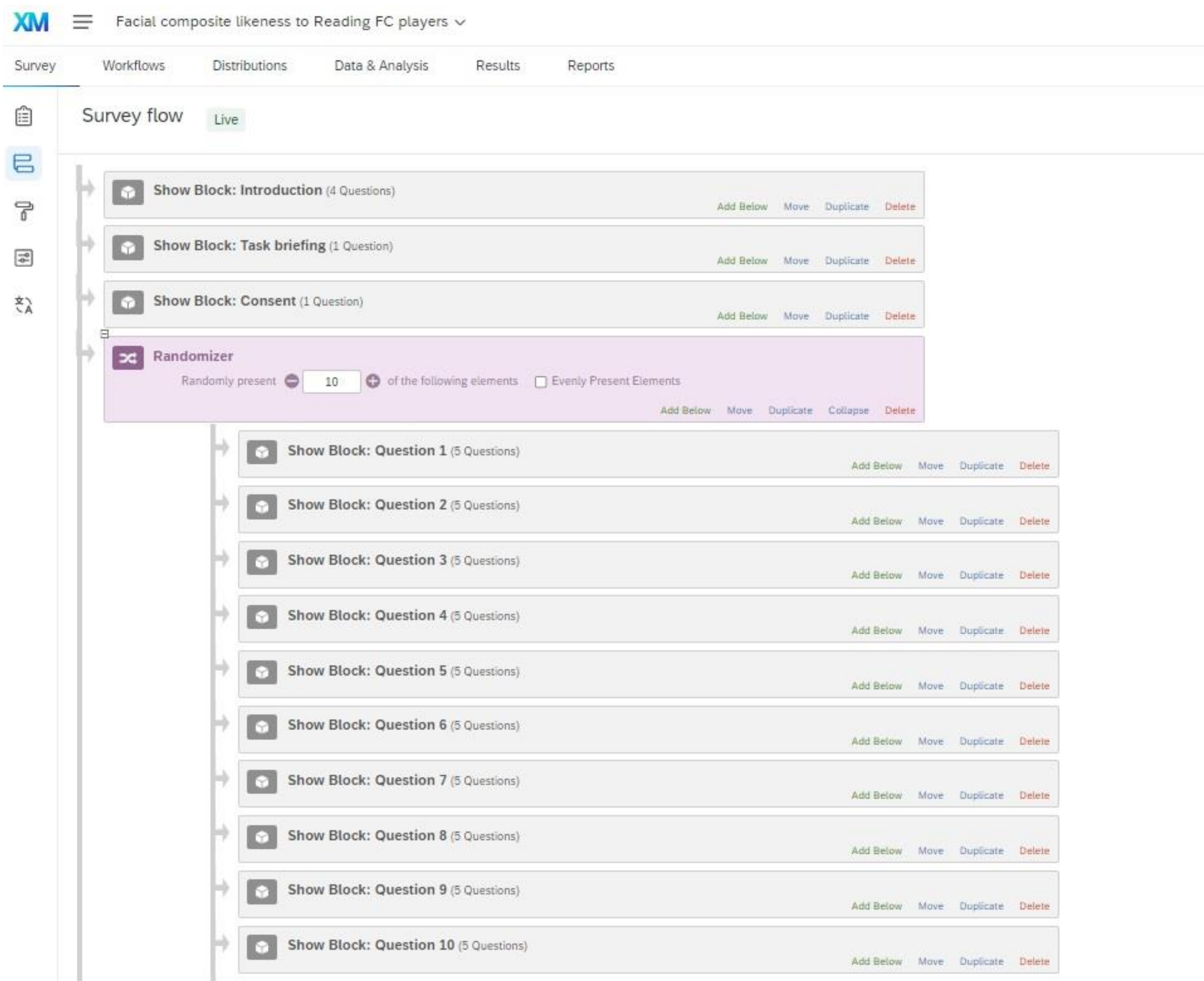


Fig 4. Settings of the target/composite set randomisation.

An example of the layout of the composites and targets in the likeness rating survey. See below:

Fig 5. Layout of the composites and the target.

Q9. Please look at all composites first and then go back and rate each in turn for overall match to target. Use the rating scale from 1=low likeness to 7=high likeness.

28

Rating for this composite:

Q9. Please look at all composites first and then go back and rate each in turn for overall match to target. Use the rating scale from 1=low likeness to 7=high likeness.

33

Rating for this composite:

Fig 6. A zoomed in view of the above example.

Information sheets for composite likeness rating, Experiment 2

Briefing sheet (likeness rating)

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites and has received ethical approval.

If you agree to take part, your involvement will be voluntary and will last for about 15 minutes. You will be presented with composites of football players who compete at an international level in the UK. I will ask you to give a rating of likeness for each composite image in the presence of the relevant target photograph. For each image, I'll ask you to give a rating from 1 (very-poor likeness) to 7 (very-good likeness).

Other people will also be asked to rate the composites in the same way. When this is done, the plan is to pool these data and analyse them using statistics to see if the experimental aims have been met (these will be explained to you at the end of today's session). In that analysis, your data will not be identifiable as it is anonymous; the analysis will also be based on group (not individual) data. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Any point during your involvement in the experiment, you have the right to withdraw and not complete the experiment. Just tell me this at any time; any data collected will then not be used. Please note that at the end of today's research session with you, it will not be possible to remove this data since it will not have your name associated with it and cannot be identified when mixed in with other people's data.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries, and concerns, please feel free to contact me or my supervisor using the following details.

Thank you,

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Debriefing sheet (likeness rating)

Please keep this page for your information.

My name is Heidi Kuivaniemi-Smith, a PhD student under the supervision of Dr Charlie Frowd from the Department of Psychology at the University of Winchester. I am carrying out a project in the area of facial composites, which are pictures of faces constructed from a person's memory. The study is developing procedures for producing identifiable sketch-based composites and has received ethical approval.

In the study, you looked at sketch composites that were constructed by previous participants in one of four conditions. In the control condition, composites were constructed in a similar way to eyewitnesses, by describing the target face using cognitive interviewing techniques and then guiding the artist to create a sketch of the target. In an alternative condition, participants used the PRO-fit composite system to identify facial features for sketching. For all participants, the exposure to the target face was 5 or 60 seconds.

Your likeness-rating data will be mixed anonymously with data from other participants in the study. These data will then be analysed to see if the aims of the experiment have been met.

As a result of this analysis, it is expected that composites will have higher likeness ratings when constructors used PRO-fit composite system to select facial features in the context of a complete face after which a sketch was created. This is expected in both target exposure conditions but more of a benefit is expected for the 5 second exposure time. Ultimately, the aim is to improve the sketching method to produce more identifiable composites. We would also be looking to publish results from group data in scientific journals and discuss them at academic conferences.

Just to remind you that you have the right to have your data withdrawn and thus not used for the following analysis stage of the experiment. Please tell me now if you would like to do this. Otherwise, your data will be mixed in anonymously with other people's data and so it will not be possible to identify it for retraction.

Thank you for participating. For further information on the topic or other information, queries, and concerns, please feel free to contact me or my supervisor using the following details.

If you have concerns about the conduct of the research, would you please contact Dr Louise Bunce, Chair of Psychology Ethics, on Louise.Bunce@winchester.ac.uk or using the university address (below). For any further information on the topic or other information, queries, and concerns, please feel free to contact me or my supervisor using the following details.

Heidi Kuivaniemi-Smith. Email: H.Kuivaniemi-Smit.14@unimail.winchester.ac.uk

Mob: 07979977334

Project supervisor -

Dr Charlie Frowd, Senior Lecturer, Department of Psychology, University of Winchester, Winchester SO22 4NR

Tel: (01962) 841515. Fax: (01962) 842280. E-mail: Charlie.Frowd@winchester.ac.uk

Reference. General information about facial composites may be found in the following review chapter: Frowd (2014). Facial composite systems. In T. Valentine and J. Davis (Eds.) Forensic Facial Identification. Wiley-Blackwell. (It is available for download at http

An example of the likeness rating sheet used for data collection:

Participant number: **Age:**

Please rate the overall likeness of each composite sketch to the target image on a scale of **1-7**

(1 – poor likeness, 7 – good likeness).

Target 1:

Composite 28:

Composite 39:

Composite 30:

Composite 9:

Target 2:

Composite 18:

Composite 34:

Composite 10:

Composite 22:

Target 3:

Composite 20:

Composite 3:

Composite 35:

Composite 6:

Target 4:

Target 6:

Composite 32:

Composite 21:

Composite 27:

Composite 17:

Target 7:

Composite 37:

Composite 7:

Composite 15:

Composite 16:

Target 8:

Composite 26:

Composite 29:

Composite 4:

Composite 14:

Target 9:

Composite 1:

Composite 8:

Composite 23:

Composite 11:

Composite 19:

Composite 33:

Composite 13:

Composite 38:

Target 5:

Composite 24:

Composite 40:

Composite 25:

Composite 31:

Target 10:

Composite 2:

Composite 12:

Composite 5:

Composite 36:

6.9 APPENDIX J – Merging sketching with a scientific approach: Interview with Forensic Artist

(Thom Shaw T. Shaw, personal communication, July 6, 2017)

DNA analysis has revolutionised forensic investigations in the past few decades, helping to solve cases where a misleading eyewitness account has led to an innocent suspect being convicted of a crime they did not commit (e.g., Devlin, 1976). This has not, however, helped to identify suspects whose DNA is not on a database. The latest development employing a sketch artist involves a system that can predict some facial features from a subject's DNA, and create a facial image based on that analysis. For example, the snapshot DNA phenotyping service by Parabon Nanolabs 'reads tens of thousands of genetic variants ("genotypes") from a DNA sample and uses this information to predict what an unknown person looks like' (https://snapshot.parabon-nanolabs.com/phenotyping). The service states that their forensic artists are trained to conduct cognitive interviews and produce composites from witnesses' descriptions remotely via screen sharing technology. If DNA is available, Snapshot can provide a prediction for the face and their forensic artists can combine the composite created from the eyewitness account with the likeness predicted by Snapshot. The service claim that the result is highly accurate, combining best sources of information.

The researcher interviewed Parabon's forensic artist Thom Shaw via Skype and email about his work that combines DNA information and forensic art. Thom Shaw's job is primarily to make small adjustments to Snapshot composites after being generated by the software. He is first provided with a bald head that is generated by the software. Much of the work on the face has already been done at this stage. The skin tone is set, along with facial features. Multiple texture maps are used to create the skin and features, which tends to reduce the quality in some areas such as eyes and eyebrows. The ears need altering as they are not anatomically correct. The neck does not have a texture map applied to it, which therefore lacks detail, and the height of the forehead is generally inaccurate (too low). Freckles are not included in the initial image. Shaw reviews information provided to him as a report by Parabon's Bioinformatics scientists and then enhances certain features using reference libraries, colour protocols and Photoshop, while staying within the boundaries of the predictions.

Shaw uses Photoshop and a Wacom Cintiq to outline areas of the face to be enhanced, to ensure the features remain accurate. He uses a catalogue of stock images to locate eyes that meet the criteria for shape and skin tone. He copies one eye, pastes it into a new layer on the composite, and blends and adjusts it to match the prediction for shape and skin tone. DNA does not predict asymmetry in a face, and so an eye is copied and mirrored to the other side. An iris is chosen from an approved set to match the prediction and pasted onto both eyes. The same process is used to select eyebrows, which are blended to match the predicted hair colour and shape. Ear shape cannot be predicted from the DNA yet, and so a generic feature is chosen and blended to match the skin tone. Next, a neck with a similar skin tone is copied and pasted onto the image and blended to fit the head. A consistency chart and colour spectra is utilised to select an appropriate colour for the hair. A generic hairstyle is chosen from a reference catalogue and copied onto the composite and blended to fit the head. Freckles are added according to the prediction. The range is zero, few, some or many. Shaw notes that if investigators have specific information about a particular suspect which needs to be incorporated into the composite, there is a form on which this can be requested.

Most of Shaw's work consists of other forensic artwork such as age progression and 2D facial reconstruction using skull and phenotype predictions, rather than for facial composites. It appears that the approach with composites is very technical and requires good digital artistic skills. It is unclear if the initial DNA composite can be tweaked from an eyewitness's description during a cognitive interview or whether this occurs only from the information provided by the investigating officers. In any case, forensic artists would need to be comfortable with their digital skills to do this, and the technical aspect might mean that a witness is involved less in the process, conceivably leaving him or her feeling redundant for most of the interview. A composite that lacks eyewitness account risks becoming generic looking at least for some parts such as hair and ears, both areas of the face that can provide important information for identification, especially if distinctive. Note that, since DNA cannot predict asymmetry, important information about facial configuration and / or shapes of features may be missing in the resulting face.

7 REFERENCES:

Almeida, M., Challa, M., Ribeiro, M., Harrison, A.M. and Castro, M.C. (2022), Editorial Perspective: The mental health impact of school closures during the COVID-19 pandemic. *Journal of Child Psychology and Psychiatry*, 63: 608-612.

Aloisi, A. & De Stefano, V. (2022). Essential jobs, remote work and digital surveillance: Addressing the COVID-19 pandemic panopticon. *International Labour Review*, *161*(2), 289–314. https://doi.org/10.1111/ilr.12219

Arnold J. G. (1982) *Police artist versus identikit survey results.* Report to the National Research Council of Canada. File reference no.4050–1.

Association of Chief Police Officers (Scotland), National Policing Improvement Agency (2009). *Facial identification guidance*. http://library.college.police.uk/docs/acpo/facial-identification-guidance-2009.pdf

Baddeley, A. D., & Woodhead, M. (1983). "Depth of processing, context, and face recognition": Erratum. *Canadian Journal of Psychology / Revue canadienne de psychologie, 37*(2). https://doi.org/10.1037/h0080889

Bahrick, H. P. (1969). Discriminative and associative aspects of pictorial paired-associate learning: Acquisition and retention. *Journal of Experimental Psychology, 80*(1), 113–119. https://doi.org/10.1037/h0027122

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of experimental psychology: General*, *104*(1), 54.

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist*, *44*(9), 1185.

Berman, G. L., & Cutler, B. L. (1998). The influence of processing instructions at encoding and retrieval on face recognition accuracy. *Psychology, Crime and Law*, *4*(2), 89-106.

Bilodeau, E. A. (1967) Experimental interference with primary associates and their subsequent recovery with rest. *Journal of Experimental Psychology*, 73, 328-332.

Blanz, V., Albrecht, I., Haber, J., & Seidel, H. P. (2006). Creating Face Models from Vague Mental Images. *Computer Graphics Forum*, *25*, 645-654.

Bond, N.W., & McConkey, K.M. (1995). Information retrieval: Reconstructing faces. In N. Brewer & C. Wilson (Eds.), *Psychology and policing* (pp. 101-117). Erlbaum.

Bower, B. (1990). Gone but not forgotten: scientists uncover pervasive, unconscious influences on memory. *The Free Library*, https://www.thefreelibrary.com/Gone but not forgotten: scientists uncover pervasive, unconscious...-a09125984

Bower, G.H. (1981). Mood and Memory. *American Psychologist*, 36(2), 129-148.

Boylan, J. (2000). *Portraits of Guilt*. Pocket Books: New York.

Bregman, A. S. (1968). Forgetting curves with semantic, phonetic, graphic, and contiguity cues. *Journal of Experimental Psychology, 78*(4, Pt.1), 539-546. https://doi.org/10.1037/h0026635

Brigham, J. C., Bennett, L. B., Meissner, C. A., & Mitchell, T. L. (2007). The influence of race on eyewitness memory. In *The Handbook of Eyewitness Psychology: Volume II* (pp. 271-296). Psychology Press.

Brown, C., Frowd, C. D., & Portch, E. (2017, September). Tell me again about the face: Using repeated interviewing techniques to improve feature-based facial composite technologies. In *2017 Seventh International Conference on Emerging Security Technologies (EST)* (pp. 38-43). IEEE.

Brown, C., Portch, E., Nelson, L., & Frowd, C. D. (2020). Reevaluating the role of verbalization of faces for composite production: Descriptions of offenders matter! *Journal of Experimental Psychology: Applied, 26*, 248–265.

Bruce, V. (1982). Changing faces: visual and non-visual coding processes in face recognition.

British Journal of Psychology, 73(1), 105-116.

Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. (1992). The Importance of 'Mass' in Line Drawings of Faces. *Applied Cognitive Psychology*, 6(7), 619-628.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*(4), 339.

Bruce, V., Ness, H., Hancock, P. J., Newman, C., & Rarity, J. (2002). Four heads are better than one: Combining face composites yields improvements in face likeness. *Journal of Applied Psychology*, *87*(5), 894.

Bruce, V., Valentine, T., & Baddeley, A. (1987). The basis of the ¾ view advantage in face recognition. *Applied Cognitive Psychology*, 1 (2), 109-120.

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

Bruce, V., & Young, A. (1986) Understanding face recognition. *British Journal of Psychology*, 77(3), 305-327.

Bruyer, R. (2011). Configural face processing: A meta-analytic survey. *Perception*, *40*(12), 1478-1490.

Bruyer, R., & Crispeels, G. (1992). Expertise in person recognition. *Bulletin of the Psychonomic Society*, *30*(6), 501-504.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243-248.

Caldwell, C. & Johnston, V.S. (1991). Tracking a criminal suspect through face-space with a genetic algorithm. *In Proceedings of the Fourth International Conference on Genetic Algorithms*, p. 416-421.

Capaldi, E. J., & Neath, I. (1995). Remembering and forgetting as context discrimination. *Learning & Memory*, *2*(3-4), 107-132.

Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 95-103.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, *195*(4275), 312-314.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children?. *Visual cognition*, *1*(2-3), 253-274.

Chance, J. E., Turner, A. L., & Goldstein, A. G. (1982). Development of differential recognition for own-and other-race faces. *The Journal of psychology*, *112*(1), 29-37.

Chen, W., & Liu, C. H. (2009). Transfer between pose and expression training in face recognition. *Vision Research*, *49*(3), 368-373.

Christie, D.F.M and Ellis, H.D. (1981). Photofit constructions versus verbal descriptions of faces. Journal of Applied Psychology, 66(3), 358-363.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.

Clark, S. E., Moreland, M. B., & Rush, R. A. (2015). Lineup composition and lineup fairness. *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV*, 127-157.

Clifford and George (1996). A field evaluation of training in three methods of witness/victim investigative interviewing. Psychology, Crime and Law, 2(3), 231-248.

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology,* 17(1), 97-116.

Cohen, G., & Conway, M. A. (Eds.). (2007). *Memory in the real world*. Psychology press.

Cole, P., & Pringle, P. (1974). *Can you positively identify this man?* London: Andre Deutsch.

Comish, S. E. (1987). Recognition of facial stimuli following an intervening task involving the Identi-kit. *Journal of Applied Psychology*, 72(3), 488-491.

Conway, M. A. (1993). Method and meaning in memory research. In G. M. Davies & R. H. Logie (Eds.), Memory in everyday life (pp. 499-524). North-Holland: Elsevier Science Publishers.

Craw, I., & Cameron, P. (1991, September). Parameterising images for recognition and reconstruction. In *BMVC91: Proceedings of the British Machine Vision Conference, organised for the British Machine Vision Association by the Turing Institute 24–26 September 1991 University of Glasgow* (pp. 367-370). Springer London.

Cutler, B.L. & Penrod, S.D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology*, 73(2), 281.

Cutler, B.L., Penrod, S.D. & Martens, T.K. (1987b). Improving the reliability of eyewitness identifications: 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 computerized facial composite production system. *Forensic Reports*.

Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology*, *32*(4), 331-331.

Dando, C.J., Ormerod, T.C., Wilcock, R. & Milne, R. (2011). When help becomes hindrance: Unexpected errors of omission and commission in eyewitness memory resulting from change temporal order at retrieval? *Cognition*, 121 (3), 416-421.

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.

Darnbrough, M. (1977). The use of facial reconstructions methods by the police. In *Annual Conference of the British Psychological Society, Exeter, Devon*.

Davies, G. (1986). Capturing likeness in eyewitness composites: The police artist and his rivals. *Medicine, Science and the Law*, *26*(4), 283-290.

Davies, G. M. (1983). Forensic face recall: the role of visual and verbal information. In S. M. A. Lloyd-Bostock and B.R. Clifford (Eds.), *Evaluating Witness Evidence* (pp.103-123). Chichester: John Wiley and Sons.

Davies, G. M., & Christie, D. (1982). Face recall: an examination of some factors limiting composite production accuracy. *Journal of Applied Psychology*, *67*, 103-109.

Davies, G. M., Ellis, H. D., & Shepherd, J. W. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, 63(2), 180-187.

Davies, G. M., Ellis, H., G., & Shepherd, J. W. (1978). Face identification: The influence of delay upon accuracy of photofit construction. *Journal of Police Science and Administration*, 6, 35-42.

Davis, J. P., Gibson, S., & Solomon, C. (2014). The positive influence of creating a holistic facial composite on video line-up identification. *Applied Cognitive Psychology*, *28*(5), 634-639.

Davies, G. & Milne, A. (1985). Eyewitness composite production: A function of mental and physical reinstatement of context. *Criminal Justice and Behavior*, 12(2), 209-220.

Davies, G.M., & Milne, A. (1982). Recognizing Faces in and out of Context. *Current Psychological Research*, 2(4), 235-246.

Davies, G., Milne, A. & Shepherd, J. (1983). Searching for operator skills in face composite reproduction. *Journal of Police Science and Administration*, 11 (4), 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., Van der Willik, P., & Morrison, L. J. (2000). Facial composite production: A comparison of mechanical and computer-driven systems. *Journal of Applied Psychology*, *85*(1), 119.

Department of Justice. (2015). *Pace Code D: Code of Practice for the identification of persons by police officers*. https://www.justice-ni.gov.uk/publications/pace-code-d-identification-persons-by-police-officers

Devlin, Hon. Lord P. (1976). *Report to the Secretary of state for the Home department of the departmental committee on evidence of identification in criminal cases*. London: Her Majesty's Stationery Office.

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(2), 129-139.

Ellis, H. D., & Flin, R. H. (1990). Encoding and storage effects in 7-year-olds' and 10-year-olds' memory for faces. *British Journal of Developmental Psychology*, *8*, 77-92.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*(4), 431-439.

Ellis, H. 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(1), 101-106.

Ellis, H. D., Davies, G. M., & Shepherd, J. W. (1978). A critical examination of the Photofit system for recalling faces, *Ergonomics*, *21*, 4, 297-307.

Ellis, H. D., & Shepherd, J. W. (1992). Face memory: Theory and practice. *Aspects of memory*, *1*, 136-143.

Ellis, H., Shepherd, J., & Davies, G. (1975). AN INVESTIGATION OF THE USE OF THE PHOTO-FIT* TECHNIQUE FOR RECALLING FACES. *British Journal of Psychology*, *66*(1), 29-37.

Erickson, W. B., Brown, C., Portch E., Lampinen, J. M., Marsh, J. M., Fodarella, C., Petkovic, A., Coultas, C., Newby, A., Data, L., Hancock, P. J. B., & Frowd, C. D. (2022). The impact of weapons and unusual objects on the construction of facial composites, *Psychology, Crime & Law*, DOI: 10.1080/1068316X.2022.2079643

Eysenck, M. W., & Keane, M. T. (2015). *Cognitive psychology: A student'handbook*. Psychology press.

Farah, M. (1991). Patterns of Co-occurrence Among the Associative Agnosias: Implications for Visual Object. *Human Cognitive Neuropsychology: A Textbook with Readings*, 367-394.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 628.

Faundez-Salinas, G. (2017). *An investigation into the influence of varying the witness' visual field of the artist/interviewer's space of work, when producing facial composite sketches using cognitive interviews via webcam.* Masters Dissertation, University of Dundee.

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

Fisher, R. P., & Geiselman, R. E. (2010). The cognitive interview method of conducting police interviews: Eliciting extensive information and promoting therapeutic jurisprudence. *International journal of law and psychiatry*, *33*(5-6), 321-328.

Fisher, R. P., Geiselman, R. E., Raymond, D. S. & Jurkevich, L.M., et al. (1987). Enhancing enhanced eyewitness memory: Refining the cognitive interview. *Journal of Police Science & Administration,* 15(4), 291-297.

Flexser, A. J., & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review*, *85*(3), 153.

Fodarella, C., Frowd, C. D., Warwick, K., Hepton, G., Stone, K., Date, L. & Heard, P. (2017). Adjusting the focus of attention: helping witnesses to evolve a more identifiable composite. Forensic Research & Criminology International Journal. 5(1), 221-230.

Fodarella, C., Kuivaniemi-Smith, H., Gawrylowicz, J., & Frowd, C. D. (2015). Forensic procedures for facial-composite construction. *Journal of Forensic Practice,* 17(4), 259 – 270.

Fodarella, C., Marsh, J. E., Chu, S., Athwal-Kooner, P., Jones, H. S., Skelton, F. C., ... & Frowd, C. D. (2021). The importance of detailed context reinstatement for the production of identifiable composite faces from memory. *Visual Cognition*, *29*(3), 180-200.

Frowd, C. D. (2017) *Facial composite systems: Production of an identifiable face.* In: Face Processing: Systems, Disorders and Cultural Differences. Nova Science Publishers.

Frowd, C.D. (2015). Facial composites and techniques to improve image recognisability. In T. Valentine, & J. Davis (Eds.) *Forensic facial identification: theory and practice of identification from eyewitnesses, composites and cctv* (pp. 43 - 70.) Chichester:Wiley-Blackwell.

Frowd, C. D. (2012). Facial Recall and Computer Composites. In C. Wilkinson and C. Rynn (Eds) *Facial Identification*. Cambridge: University Press. Cambridge.

Frowd, C. D. (2021). Forensic Facial Composites. In Toglia, M., Smith, A., & Lampinen, J. M. (Eds.) *Methods, Measures, and Theories in Forensic Facial-Recognition* Taylor and Francis (pp. 34-64). Taylor and Francis.

Frowd, C. D., Bruce, V., McIntyre, A. H., Ross, D., Fields, S., Plenderleith, Y., & Hancock, P. J. (2006). Implementing holistic dimensions for a facial composite system. *Journal of Multimedia*, *1*(3), 42-51.

Frowd, C.D., Bruce, V., Ness, H., Bowie, L., Thomson-Bogner, C., Paterson, J., McIntyre, A., & Hancock, P.J.B. (2007). Parallel approaches to composite production. *Ergonomics*, *50*, 562-585.

Frowd, C.D., Bruce, V., Ross, D., McIntyre, A., & Hancock, P. J. B. (2007). An application of caricature: How to improve the recognition of facial composites. *Visual Cognition*, 15(8), 954–984. https://doi.org/10.1080/13506280601058951

Frowd, C. D., Bruce, V., Smith, A., & Hancock, P. J. B. (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied*, 14(3), 276-287.

Frowd, C. D., Carson, D., Ness, H., McQuiston, D., Richardson, J., Baldwin, H., & Hancock, P. J. B. (2005a). Contemporary Composite Techniques: the impact of a forensically-relevant target delay. *Legal & Criminological Psychology*, 10(1), 63-81.

Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S., & Hancock, P. J. B. (2005b). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law*, 11, 33-52.

Frowd, C. D., Erickson, W. B., Lampinen, J. L., Skelton, F. C., McIntyre, A. H., & Hancock, P. J. B. (2015). A Decade of Evolving Composites: Regression- and Meta-Analysis. *Journal of Forensic Practice*,17(4), 319-334.

Frowd, C. D., & Fields, S. (2011). Verbalization effects in facial composite production. *Psychology, Crime & Law*, *17*(8), 731-744.

Frowd, C. D. & Goodfellow, Y. (2018). An investigation into whether poor quality composites are a result of diminished facial recall. In M. Bindemann and A. Megreya (Eds.) *Face Processing: Systems, Disorders and Cultural Differences.* Nova Science.

Frowd, C. D., Hancock, P. J. B., Bruce, V., Skelton, F. C., Atherton, C., Nelson, L., *et al.* (2011a). Catching more offenders with EvoFIT facial composites: lab research and police field trials. *Global Journal of Human Social Science*, 11(3), 46-58.

Frowd, C.D., Hancock, P.J., & Carson, D. (2004). EvoFit: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on applied perception,* 1(1), 19-39.

Frowd, C. D., Nelson, L., Skelton F. C., Noyce, R., Heard, P., Henry, J. *et al*. (2012d). Interviewing techniques for Darwinian facial composite systems. *Applied Cognitive Psychology*, 26(4), 576-584.

Frowd, C., Park, J., McIntyre, A., Bruce, V., Pitchford, M., Fields, S., ... & Hancock, P. J. (2008, August). Effecting an improvement to the fitness function. How to evolve a more identifiable face. In *Bio-inspired Learning and Intelligent Systems for Security, 2008. BLISS'08. ECSIS Symposium on* (pp. 3-10). IEEE.

Frowd, C. D., Portch, E., Killeen, A., Mullen, L., Martin, A. J., & Hancock, P. J. (2019, July). EvoFIT facial composite images: a detailed assessment of impact on forensic practitioners, police investigators, victims, witnesses, offenders and the media. In *2019 eighth international conference on emerging security technologies (EST)* (pp. 1-7). IEEE.

Frowd, C. D., Skelton, F. C., Atherton, C., Pitchford, M., Bruce, V., Atkins, R., Gannon, C., Ross, D., Young, F., Nelson, L., Hepton, G., McIntyre, A. H., & Hancock, P. J. B. (2012). Understanding the multi-frame caricature advantage for recognising facial composites. *Visual Cognition*, 20, 1215-1241.

Frowd, C. D., Skelton, F. C., Butt, N., Hassan, A., & Fields, S. (2011). Familiarity effects in the construction of facial-composite images using modern software systems. *Ergonomics*, *54*, 1147-1158.

Frowd, C. D., Skelton F., Hepton, G., Holden, L., Minahil, S., Pitchford, M., McIntyre, A., Brown, C. D., & Hancock, P. J. B. (2013). Whole-face procedures for recovering facial images from memory. *Science & Justice*, 53(2), 89-97.

Fysh, M. C. (2018). Individual differences in the detection, matching and memory of faces. *Cognitive Research: Principles and Implications*, *3*(1), 1-12.

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior*, *33*(4), 298-307.

Gambell, S. B. (2006). The Need to Revisit the Neil v. Biggers Factors: Suppressing Unreliable Eyewitness Identification. *Wyoming Law Review*, *6*(1), 189–221. https://doi.org/10.59643/1942-9916.1126

Gauthier, I., McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & VanGulick, A. E. (2014). Experience moderates overlap between object and face recognition, suggesting a common ability. *Journal of Vision*, 14 (7).

Geiselman, R. E., Fisher, R. P., Cohen, G., & Holland, H. L. (1986). Eyewitness responses to leading and misleading questions under the cognitive interview. *Journal of Police Science and Administration*, *14*, 31-39.

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 the cognitive interview. *Journal of Police Science & Administration, 12*(1), 74-80.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the cognitive interview. *The American journal of psychology*, 385-401.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, 70, 401-412.

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. In *Proceedings of the 10th annual conference on Genetic and evolutionary computation* (pp. 1485-1490).

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

Gibson, L. (2010). *Forensic art essentials: a manual for law enforcement artists*. Academic Press.

Gibson, S. J., Solomon, C. J., Maylin, M. I. S., & Clark, C. (2009). New methodology in facial composite construction: from theory to practice, *International Journal of Electronic Security and Digital Forensics*, 2(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 and J.A. Bangham (eds), Proceedings of the British Machine Vision Conference 2003 1: 221-230.

Gillenson, M. L., & Chandrasekaran, B. (1975, June). WHATSISFACE: human facial composition by computer graphics. In *Proceedings of the 2nd annual conference on Computer graphics and interactive techniques* (pp. 212-221).

Glenberg, A. M. (1997). What memory is for. *Behavioral and brain sciences*, *20*(1), 1-19.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology*, *66*(3), 325-331.

Goffaux, V.& Rossion, B. (2006). Faces are ‘spatial’ –holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4), 1023.

Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. *The journal of psychology*, *105*(1), 47-59.

Hancock, P. J. (2000). Evolving faces from principal components. *Behavior Research Methods, Instruments, & Computers*, *32*(2), 327-333.

Green, D. L., & Geiselman, R. E. (1989). Building composite facial images: Effects of feature saliency and delay of construction. *Journal of Applied Psychology*, *74*(5), 714.

Hammond, L., Wagstaff, G. F., & Cole, J. (2006). Facilitating eyewitness memory in adults and children with context reinstatement and focused meditation. *Journal of Investigative Psychology and Offender Profiling*, *3*(2), 117-130.

Hancock, P. J., Burke, K., & Frowd, C. D. (2011). Testing facial composite construction under witness stress. *International Journal of Bio-Science and Bio-Technology*, *3*(3), 65-71.

Hancock, P. J., Burton, A. M., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory & cognition*, *24*, 26-40.

Hancock, P. J., & Frowd, C. D. (2002). Evolutionary generation of faces. In *Creative evolutionary systems* (pp. 409-423). Morgan Kaufmann.

Hancock, P. J., Wilkinson, C., & Rynn, C. (2012). Unfamiliar face recognition. *Craniofacial Identification*, 11-23.

Hay, D. C., Young, A. W. (1982) “The human face” in *Normality and Pathology in Cognitive Functions* ed Ellis A. W. (London: Academic Press) pp 173–202.

Hayward, W. G., Rhodes, G., Winkler, C., & Schwaninger, A. (2005). Own-race face effects in processing of configural and component information by Chinese observers. *Journal of Vision*, *5*(8), 534-534.

Hockley, W. E. (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1412.

Hockley, W. E., Bancroft, T. D., & Bryant, E. (2012). Associative and familiarity-based effects of environmental context on memory. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale*, 66(2), 81–89.

Hole, G., & Bourne, V. (2010). *Face processing: Psychological, neuropsychological, and applied perspectives*. Oxford University Press.

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, *31*(10), 1221-1240.

Homa, G. (1983). The law enforcement composite sketch artist. *Berlin, NJ: Community*.

Huang, J., Heisele, B., & Blanz, V. (2003). Component-based face recognition with 3D morphable models. In *Audio-and Video-Based Biometric Person Authentication: 4th International Conference, AVBPA 2003 Guildford, UK, June 9–11, 2003 Proceedings 4* (pp. 27-34). Springer Berlin Heidelberg.

Jackson, E. (2023). *Optimising EvoFIT by reducing the number of faces shown during composite construction*. PhD Thesis, University of Central Lancashire.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313-323.

Johnson, G. (2010). *Masters Dissertation*, University of Dundee.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577-596.

Joseph, R. M., & Tanaka, J. (2003). Holistic and part-based face recognition in children with autism. *Journal of child psychology and psychiatry*, *44*(4), 529-542.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, *17*(11), 4302-4311.

Kaufmann, J. M., & Schweinberger, S. R. (2008). Distortions in the brain? ERP effects of caricaturing familiar and unfamiliar faces. *Brain research*, *1228*, 177-188.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues. *Perception*, *25*(1), 37-52.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *11*(3), 211-222.

Kitson, T., Darnbrough, M., A. Shields, E. "Lets face it." Police Research Bulletin, 1978, No. 30, 7-13.

Klare, B., Li, Z & Jain, A. (2011). Matching Forensic Sketches to Mug Shot Photos, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33 (3), p.639-646.

Klatzky, R. L., & Forrest, F. H. (1984). Recognizing familiar and unfamiliar faces. *Memory & cognition*, *12*, 60-70.

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

Kokila, R., Sannidhan, M. S., & Bhandary, A. (2017, September). A novel approach for matching composite sketches to mugshot photos using the fusion of SIFT and SURF feature descriptor. In *2017 international conference on advances in computing, communications and informatics (ICACCI)* (pp. 1458-1464). IEEE.

Kovera, M. B., Penrod, S. D., Pappas, C., & Thill, D. L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology*, *82*(2), 235.

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

Krouse, F. L. (1981) ‘Effects of pose, pose change, and delay on face recognition performance’, Journal of applied psychology, 66(5), pp. 651–654. doi: 10.1037/0021-9010.66.5.651.

Kuivaniemi-Smith, H. J., Frowd, C. D., Sanderson, H. & Minahil, S. (2013) Improving the effectiveness of sketch-based composite images [Unpublished manuscript]. Department of Psychology, University of Winchester.

Kuivaniemi-Smith.H. J., Nash, R. A., Brodie, E. R., Mahoney, G., & Rynn, C. (2014). Producing facial composite sketches in remote Cognitive Interviews: a preliminary investigation. *Psychology, Crime & Law,* 20(4), 389-406.

Kvavilashvili, L., & Ellis, J. (2004). Ecological validity and twenty years of real-life/laboratory controversy in memory research: A critical (and historical) review. *History and Philosophy of Psychology*.

Köhnken, G., Milne, R., Memon, A. & Bull, R. (1999) The cognitive interview: A meta-analysis, Psychology, Crime & Law, 5:1-2, 3-27, DOI: 10.1080/10683169908414991

Lampinen, J. L., Erickson, B., Frowd, C. D., & Mahoney, G. (2015). Mighty morphin’ age progressions: how artist, age range, and morphing influences the similarity of age progressions to target individuals. *Psychology, Crime & Law*, 21, 952-967.

Lander, K. & Bruce, V. In Otani, H., & Schwartz, B.L. (Eds.). (2018). *Handbook of Research Methods in Human Memory* (1st ed.). Routledge. https://doi.org/10.4324/9780429439957

Laughery, K. R., Alexander, J. F., & Lane, A. B. (1971). Recognition of human faces: effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology*, *55*(5), 477.

Laughery, K. R., Duval, G. C., & Fowler, R. H. (1977). An analysis of procedures for generating facial images. *Mug File Project, Report No. UHMUG-2. Houston, TX: University of Houston*.

Leijtens, M.A.L.A. (2016) *Recognition of Facial Composites: a Comparison Between PRO-fit and Forensic Sketches in Relation to Target Viewing Time*. [Masters Thesis, University of Amsterdam].

Light, L. L., Kayra Stuart, F. & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, 5(3), 212-228.

Loftus, E. F., & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist*, *35*(5), 409.

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about “weapon focus”. *Law and human behavior*, *11*(1), 55-62.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, *13*(5), 585-589.

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*(1), 34-39.

MacDonald, S. (1984). Wanted art: Meet a man who draws shady characters. *The Wall Street Journal*, 1.

MacLin, O. H., MacLin, M. K., & Malpass, R. S. (2001). Race, arousal, attention, exposure and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, and Law*, *7*(1), 134.

Malpass, R.S., & Devine, P.G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. Journal of Applied Psychology, 66(4), 482.

Malpass, R.S., & Lindsay, R.C. (1999). Measuring lineup fairness. Applied Cognitive Psychology, 13(Suppl 1.), S1-S8.

Mancusi, S. (2010). *The police composite sketch*. Springer Science & Business Media.

Marsh, J. E., Demaine, J., Bell, R., Skelton, F. C., Frowd, C. D., Röer, J. P., & Buchner, A. (2015). The impact of irrelevant auditory facial descriptions on memory for target faces: implications for eyewitness memory. *Journal of Forensic Practice*, *17*(4), 271-280.

Martin, A. J., Hancock, P. J., & Frowd, C. D. (2017, September). Breathe, relax and remember: An investigation into how focused breathing can improve identification of EvoFIT facial composites. In *2017 Seventh International Conference on Emerging Security Technologies (EST)* (pp. 79-84). IEEE.

Matzel, L. D., Schachtman, T. R., & Miller, R. R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning and Motivation*, *16*(4), 398-412.

Mauldin, M. A., & Laughery, K. R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology,* 66(3), 351-357.

McCauley, M. R., & Fisher, R. P. (1995). Facilitating children's eyewitness recall with the revised cognitive interview. *Journal of Applied Psychology*, *80*(4), 510.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, *11*, 8-15.

McQuiston-Surrett, D., Topp, L., Malpass, R. (2006) Use of facial composite systems in US law enforcement agencies. *Psychology, Crime & Law*, 12(5), 505-517.

Meinhardt-Injac, B. (2013). The context congruency effect is face specific. *Acta Psychologica*, *142*, 265-272.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. *Psychology, Public Policy, and Law*, 7(1), 3-35.

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

Memon, A., & Bull, R. (1991). The cognitive interview: Its origins, empirical support, evaluation and practical implications. *Journal of Community & Applied Social Psychology*, *1*(4), 291-307.

Memon, A., Havard, C., Clifford, B., Gabbert, F., & Watt, M. (2011). A field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. *Psychology, Crime & Law*, *17*(8), 711-729.

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology*, 94(3), 339-354.

Ministry of Justice & National Police Chiefs' Council (2022). Achieving Best Evidence in Criminal Proceedings: Guidance on Interviewing Victims and Witnesses, and Guidance on Using Special Measures. https://www.gov.uk/government/publications/achieving-best-evidence-in-criminal-proceedings

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, *31*(5), 553-566.

Murry, N. (2018). *Digital Forensic Art Techniques: A Professional's Guide to Corel Painter.* CRC Press.

Ness, H., & Bruce, V. (2006). An investigation into the use of CCTV footage to improve likeness in facial composites.

Ness, H., Bruce, V. & Hancock. (2004). Combining facial composites and CCTV. Paper presented at the 14th European Conference on Psychology and Law, Krakow, July 2004.

Ness, H., Hancock, P. J., Bowie, L., Bruce, V., & Pike, G. (2015). Are two views better than one? Investigating three-quarter view facial composites. *Journal of Forensic Practice*.

NSLEC. (2004). Practical guide to investigative interviewing. Wybosten: National Centre for Policing Excellence.

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception*, 30(6), 755-764.

Orne, M., Soskis, D., Dinges, D., & Carota Orne, E. (1984). Hypnotically induced testimony. In G. Wells & E. Lofus (Eds.), Eyewitness testimony: Psychological perspectives (pp. 171-213). New York: Cambridge University Press.

O'Toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision research*, *38*(15-16), 2351-2363.

Paulo, R. M., Albuquerque, P. B., & Bull, R. (2016). The enhanced cognitive interview: expressions of uncertainty, motivation and its relation with report accuracy. *Psychology, Crime & Law*, *22*(4), 366-381.

Peterson, J., Sommers, I., Baskin, D., & Johnson, D. (2010). The role and impact of forensic evidence in the criminal justice process (Research report submitted to the U.S. Department of Justice). Retrieved from https://www.ncjrs.gov/pdffiles1 /nij/grants/231977.pdf

Pike, G. E., Brace, N. A., Turner, J., & Vredeveldt, A. (2019). The effect of facial composite construction on eyewitness identification accuracy in an ecologically valid paradigm. *Criminal Justice and Behavior*, *46*(2), 319-336.

Police Scotland. (2021). Identification Procedures: National Guidance. https://www.scotland.police.uk/access-to-information/policies-and-procedures/guidance-documents/guidance-documents-g-i/

Postman, L., Adams, P. A., & Phillips, L. W. (1955). Studies in incidental learning: II. The effects of association value and of the method of testing. *Journal of Experimental Psychology*, *49*(1), 1.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of neuroscience*, *18*(6), 2188-2199.

Quas, J., Schaaf, JM., Alexander, K.W., Goodman, G.S. (2000) Do you really remember it happening or do you only remember being asked about it happening? Children's source monitoring in forensic contexts. In K.P., Roberts & M., Blades (Eds). *Children's source monitoring* (197-226). Lawrence Erlbaum associates.

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

Read J. D. (1995). The availability heuristic in person identification – sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology,* 9(2), 91-121.

Reiser, M. (1989). Investigative hypnosis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 151–190). Springer Publishing Company.

Reynolds, J. K., & Pezdek, K. (1992). Face recognition memory: The effects of exposure duration and encoding instruction. *Applied Cognitive Psychology*, *6*(4), 279-292.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive psychology*, *19*(4), 473-497.

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British journal of psychology*, *80*(3), 313-331.

Rhodes, G., Hayward, W. G., & Winkler, C. (2005). Expert face coding: Configural and component coding of own-race and other -race faces. *Psychonomic Bulletin and Review*, 13, 499-505. doi: 10.3758/BF03193876

Richardson, J., Stahl, D., Deal, M. et al (2010) Standards and guidelines for forensic art and facial identification. 1st ed. [PDF] p.5. Available at https://www.theiai.org/disciplines/art/ForensicArtGuidelinesSGFAFI1stEd.pdf [Accessed 15 Jun. 2016]

Robertshaw, L. (2020). The Impact of Prompted Recall on the Accuracy of Composite Sketches based on Cognitive Interview. *Master's dissertation, University of Dundee.*

Robinson, J.C. (2010). Mixed-race ancestry and its impact on the facial reconstruction process. *LSU Master's Thesis. 308*. https://doi.org/10.31390/GRADSCHOOL_THESES.308

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic bulletin & review*, *16*(2), 252-257.

Sanders, G.L. (1984). The effects of context cues on eyewitness identification responses. Journal of Applied Social Psychology, 14, 386-397.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, *16*(6), 440-444.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive psychology*, *22*(1), 36-71.

Schooler, J.W., Fiore, S.M., Brandimonte, M.A. (1997). At a loss from words: verbal overshadowing of perceptual memories. In *The Psychology of Learning and Motivation*, Medin DL (ed.). Academic Press: San Diego, CA; 293–334.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial–relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 904.

Sekuler, R., Watamaniuk, S. N., & Blake, R. (2002). Perception of visual motion.

Sergent, J. (1986). Microgenesis of face perception. *Aspects of face processing*, 17-33.

Shapiro, P. N. & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, 100(2), 139-156.

Shepherd, J. W. (1986). An interactive computer system for retrieving faces. In H. D. Ellis, M. A. Jeeves, F. Newcombe and A. Young (Eds.), *Aspects of face processing* (pp. 398–409). Martinus Nijhoff.

Shepherd, J. W., & Ellis, H. D. (1996). Face recall-methods and problems. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 87-116). Erlbaum.

Shepherd, J. W., Gibling, F. & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. *European Journal of Cognitive Psychology*, 3(1), 137-145.

Shepherd, J. W., Ellis, H. D., McMurran, M., & Davies, G. M. (1978). Effect of character attribution on Photofit construction of a face. *European journal of social psychology*.

Shepherd, J. W., Ellis, H. D., & Davies, G. M. (1982). *Identification evidence: A psychological evaluation* (pp. 71-73). Aberdeen: Aberdeen University Press.

Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological review*, *94*(4), 439.

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Josa a*, *4*(3), 519-524.

Skelton, F. C., Frowd, C. D., Hancock, P. J. B., Jones, H. S., Jones, B. C., Fodarella, C., Battersby, K., & Logan, K. (2020). Constructing identifiable composite faces: the importance of cognitive alignment of interview and construction procedure. *Journal of Experimental Psychology: Applied*, *26*, 507–521.

Skelton, F.C., Frowd, C.D., & Speers, K. (2015). The benefit of context for facial-composite construction, *Journal of Forensic Practice,* 17(4), 281-290.

Smith, M. C. (1983). Hypnotic memory enhancement of witnesses: Does it work?. *Psychological Bulletin*, *94*(3), 387.

Smith, S. M. (1988). Environmental context dependent memory. In G. M. Davies, & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13-33). New York: Wiley.

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

Smith, S. M. (1994). Theoretical principles of context-dependent. *Theoretical aspects of memory*, 168-195.

Smith, E.E., & Kosslyn, S.M. (2014). *Cognitive Psychology: Mind and Brain*. Harlow: Pearson.

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

Smith, S. M., & Vela, E. (1992). Environmental context-dependent eyewitness recognition. *Applied Cognitive Psychology*, *6*(2), 125-139.

Solomon, C. J., Gibson, S. J., & Maylin, M. (2012). EFIT-V: Evolutionary algorithms and computer composites. *Craniofacial identification*, 24-41.

Sporer, S. L. (1996). Psychological aspects of person descriptions. *Psychological issues in eyewitness identification*, 53-86.

Sporer, S. L., Tredoux, C. G., Vredeveldt, A., Kempen, K., & Nortje, A. (2020). Does exposure to facial composites damage eyewitness memory? A comprehensive review. Applied Cognitive Psychology, 34(5), 1166-1179.

Steblay, N. M. (1992). Meta-analytic Review of the Weapon Focus Effect. *Law and Human Behavior*, 16(4), 413-424.

Steinberg, S. (2011). *Steinberg's Facial Identification Catalog*. Miami.

Stop, A. (2012). (Unpublished), *Production techniques for sketching*, University of Dundee.

Stuart-Parks, C. (2017). *A Gwen Marcey Collection: A Cry from the Dust, The Bones Will Speak, When Death Draws Near*. Thomas Nelson.

Surprenant, A.M., Neath, I. (2009) *Principles of memory: Essays In Cognitive Psychology*. New York, London: Taylor & Francis Group.

Tanaka, J. W. & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology,* 46(2), 225-245.

Tanaka, J. W., & Farah, M. J. (2003). The holistic representation of faces. *Perception of faces, objects, and scenes: Analytic and holistic processes*, 53-74.

Tanaka, J.W., Kay, J.B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition*, 5, 479–496.

Tanaka, J. W. & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition,* 25(5), 583-592.

Taylor, K. T. (2001). Forensic Art and Illustration. Florida: CRC Press LLC.

The National Archives. (2005). *Criminal Justice Act 2003*. Retrieved July 12, 2023, from https://www.legislation.gov.uk/ukpga/2003/44/section/115

Thomson, D. M., Robertson, S. L., & Vogt, R. (1982). Person recognition: The effect of context. *Human Learning*, 1(1), 37-54.

Tooley, V., Brigham, J. C., Maass, A., & Bothwell, R. K. (1987). Facial Recognition: Weapon Effect and Attentional Focus 1. *Journal of Applied Social Psychology*, *17*(10), 845-859.

Tredoux, C., Nunez, D., Oxtoby, O., & Prag, B. (2006). An evaluation of ID: an eigenface based construction system: reviewed article. *South African Computer Journal*, *2006*(37), 90-97.

Tulving, E. (1983). *Elements of episodic memory*. Oxford University Press.

Tulving, E. (1972). Episodic and semantic memory. *Organization of memory*, *1*(381-403), 1.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual review of psychology*, *53*(1), 1-25.

Tulving, E. (1974). Recall and recognition of semantically encoded words. *Journal of Experimental Psychology*, *102*(5), 778.

Tulving, E., & Craik, F. I. (Eds.). (2000). *The Oxford handbook of memory*. Oxford University Press.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of verbal learning and verbal behavior*, *5*(4), 381-391.

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

Tyler, R., Towler, A., Kemp, R. I., & White, D. (2023). Let's talk about faces: Identifying faces from verbal descriptions. *British Journal of Psychology*, *114*(1), 262-281.

Vaidya, C. J., Zhao, M., Desmond, J. E., & Gabrieli, J. D. (2002). Evidence for cortical encoding specificity in episodic memory: memory-induced re-activation of picture processing areas. *Neuropsychologia*, *40*(12), 2136-2143.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43*(2), 161-204.

Valentine, T., & Heaton, P. (1999). An evaluation of the fairness of police line-ups and video identifications. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *13*(S1), S59-S72.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *17*(8), 969-993.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*(6982), 557-561.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, *36*(12), 1546.

Wells, G. L., Charman, S. D., & Olson, E. A. (2005). Building face composites can harm lineup identification performance. *Journal of Experimental Psychology: Applied*, 11, 147-156.

Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory & Cognition*, *12*(4), 338-344.

Wells, G. L. & Olson, E. (2003). Eyewitness identification. Annual Review of Psychology, 54(1), 277-295.

Wickens, D.D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological review*, (77), 1-15.

Wickham, L. H., & Swift, H. (2006). Articulatory suppression attenuates the verbal overshadowing effect: A role for verbal encoding in face identification. *Applied Cognitive Psychology*, *20*(2), 157-169.

Widden, T., Frowd, C.D., Brewer, G., Rynn, C. (2017). Improving Sketch-based Face Construction.

Wilford, M. M., & Wells, G. L. (2010). Does facial processing prioritize change detection? Change blindness illustrates costs and benefits of holistic processing. *Psychological science*, *21*(11), 1611-1615.

Wilkinson, C. (2004). *Forensic facial reconstruction*. Cambridge University Press.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*(1), 10-65.

Wood, A. (2019). *Trial of Percy Lefroy Mapleton: 86 (Notable British Trials)*. Mango Books.

Wong, C. K., & Read, J. D. (2011). Positive and negative effects of physical context reinstatement on eyewitness recall and identification. *Applied Cognitive Psychology*, *25*(1), 2-11.

Women Entrepreneurs Are Optimistic And Rapidly Adapting To A Post-Covid World. (2020). *Business World (India)*.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology*, *81*(1), 141.

Young, Andrew, W. & Burton, Mike. (2018) Are we face experts? *Trends in Cognitive Sciences*, 22 (2), 100-110.

Young, A. W., Hay, D. C., McWeeny, K. H., et al. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14(6), 737-746.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, 16(6), 747-759.

Yuille, J. C. (1993). We must study forensic eyewitnesses to know about them. *American Psychologist*, *48*(5), 572.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of applied psychology*, *71*(2), 291.

Zarkadi, Wade, K. A., & Stewart, N. (2009). Creating Fair Lineups for Suspects With Distinctive Features. *Psychological Science*, *20*(12), 1448–1453. https://doi.org/10.1111/j.1467-9280.2009.02463.x

Internet sources:

Evidence-based Justice lab. (n.d.). *William Mills*. Retrieved July 13, 2023, from https://evidencebasedjustice.exeter.ac.uk/case/william-mills/

Hanover College Psychology Department. *Psychological Research On The Net*. (n.d.). Retrieved May 29, 2018, from https://psych.hanover.edu/research/exponnet.html

Identi-Kit Solutions. (n.d.). *Identi-Kit 7*. Retrieved September 1, 2022, from https://identikit.net/company/

Parabon Nanolabs. (n.d.). *The Snaphot DNA Phenotyping Service*. Retrieved April 2, 2021, from https://snapshot.parabon-nanolabs.com/phenotyping

Premier League. (n.d.). *Premier League Explained*. Retrieved November, 25, 2023, from https://www.premierleague.com/premier-league-explained

Samantha Steinberg. (n.d.). Ethnicities Catalog. Retrieved November 25, 2023, from http://samanthasteinberg.com/Steinbergs_Ethnicities_Catalog.html

SketchCop. (n.d.). *SketchCop Software App*. Retrieved June 2022, from https://sketchcop.com/sketchcop-facial-composite-system-software/

The Crown Prosecution Service. (2018). *Identification*. Retrieved August 5, 2023, from https://www.cps.gov.uk/legal-guidance/identification.

The National Archives. (2018). *Criminal Justice (Scotland) Act*. Retrieved July 12, 2023, from https://www.legislation.gov.uk/asp/2016/1/section/57

The National Archives. (2018). Police and *Criminal Evidence Act*. Retrieved July 12, 2023, from https://www.legislation.gov.uk/ukpga/1984/60/contents

Wild Training. (2021). *Wild X Fitness App*. Retrieved July, 12, 2022, from https://www.wildtraining.co.uk/wild-blog/2021/11/16/wild-x-fitness-app

Images:

Acpo. (2009). *E-FIT made in 2006* [Digital image].

https://library.college.police.uk/docs/acpo/facial-identification-guidance-2009.pdf

Candy, J. (2016). *John Swift* [Photograph]. https://commons.wikimedia.org/wiki/File:John_Swift,_Brentford_FC,_December_2015.jpg

Egghead06. (2012). *Andy Carroll* [Photograph]. https://commons.wikimedia.org/wiki/File:AndyCarrollWHU.jpg

EvoFIT. (n.d.). *Aggravated Robbery EvoFIT* [Digital image]. https://evofit.co.uk/case-studies/

Greater Manchester Police. (2009). *Photofit kit* [Photograph]. https://www.flickr.com/photos/gmpolice1/4080920708/in/photolist-2eBtX8p-7dBMxS

Kelly, L. (2021). Attempted Child Abduction Reported in Marietta [sketch]. Patch. https://patch.com/georgia/eastcobb/attempted-child-abduction-reported-marietta

Smith, I. (2010). *Perry Fenwick* [Photograph]. https://commons.wikimedia.org/wiki/File:Perry_Fenwick_(4382642269).jpg

Tsourrai, G. & Davis, J. (2020). University of Greenwich [Digital image]. https://www.superrecognisers.com/post/impact-of-efit6-construction-on-line-up-identification-accuracy

Winterman, S. (2015). *Charley Webb* [Photograph]. https://commons.wikimedia.org/wiki/File:Charley_Webb_2015.jpg

WorthPoint. (n.d.). Vintage Identi-Kit for Police Identification of Criminals [Photograph]. https://www.worthpoint.com/worthopedia/vintage-identi-kit-police-149583425

&C. (2019). *Jeroen v Koningsbrugge Houdt Wel Zijn Kleding Aan - &C*. Wikimedia Commons [Still image from a video].
https://commons.wikimedia.org/wiki/File:Koningsbrugge2019.jpg#/media/File:Koningsbrugge2019.jpg