Asking for more information: Can recognition-based enhancements (RBE) made to internal facial features produce more accurate composites?

by

Lauren Elizabeth Ellison

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy at the University of Central Lancashire.

2023

STUDENT DECLARATION FORM

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.

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.

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: Not applicable

Signature of Candidate

Type of Award **Doctor of Philosophy**

School **Psychology**

ABSTRACT

Research focused around investigative interviewing techniques demonstrate that offering witnesses multiple opportunities to retrieve information about a crime can, considerably, improve the accuracy of information recalled from memory. However, this advantageous effect has yet to be explored in the context of composite construction. Taking this gap into consideration, the primary objective of this thesis is to investigate whether providing witnesses with further opportunity to make recognition-based enhancements (RBE) to the internal facial features of a composite, leads to subsequent improvements in composite accuracy. Across four experiments, participants were presented with an unfamiliar target face (for 60 seconds) and then constructed a composite (Experiment 1, 2 and 3: 24 hours later; Experiment 4: 3-4 hours later). Participants were randomly assigned to one of four Accumulated EvoFIT Stages (**AEFS**), each offering a different level of opportunity to make RBE to the face: **AEFS A**: involved choosing the "best" face from arrays: **AEFS B**: involved **AEFS A**, plus use of Holistic Scales: **AEFS C**: involved **AEFS B**, plus use of the Shape Tool: **AEFS D**: involved **AEFS C**, plus use of Holistic Scales for internal facial features. Experiment 1, 2 and 3 revealed that composites constructed with fewer opportunities to make RBE to internal facial features, tended to have a stronger likeness towards the target face. However, Experiment 4 revealed a noteworthy finding that, when used in combination with a holistic-cognitive interview (H-CI) and a short retention interval (3-4 hours), composites constructed with a greater opportunity to make RBE produced more accurate composites. The series of experiments suggest that the incorporation of particular interviewing techniques, and the careful optimisation of timing, can substantially enhance the accuracy of facial composites. An implication drawn from these findings is that the most accurate composites were produced when a retention interval of 3-4 hours was adopted, a timeframe that may not always be achievable to replicate in a real-world setting due to police resources.

Keywords: EvoFIT, eyewitness memory, facial composites, holistic-cognitive interview, multiple retrieval opportunities, memory retrieval, retention interval.

CONTENTS PAGE

LIST OF TABLES AND ILLUSTRATIVE MATERIAL

CHAPTER 3

CHAPTER 4

CHAPTER 5

Table 26. Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Items for Internal Features.

Table 27. Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Participants for External Features.

Table 28. Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Items for External Features.

ACKNOWLEDGEMENTS

I would like to extend my utmost appreciation to my supervisor, Dr Beth Richardson, for her constant support, invaluable guidance, and profound expertise that she has provided me with throughout the entire journey of my PhD. Her mentorship has played an instrumental role in my progression and development as a researcher, equipping me with the skills and knowledge needed for life after a PhD. I will always be thankful for that.

I would like to express my deepest gratitude to Professor Charlie Frowd, for not only inspiring my passion for research, but also for providing me with the incredible opportunity to take on this project. I would also like to thank Professor Frowd for his continued investment of time, patience, and profound guidance and expertise that he has provided me with throughout the entirety of my PhD journey.

I would also like to give my heartfelt gratitude and appreciation to my Mum and Dad for endlessly supporting and believing in me, in every way possible. I will always be in debt to you both. Finally, I would like to thank my Husband, Anthony. I am so grateful for your continuous, unwavering encouragement, endless patience, and the infinite amount of strength and support that you have provided me with. Most importantly, for fuelling my determination, by always believing in me. Without the support from you three, this would not have been possible. My heart is filled with gratitude and love for you all.

I wish to dedicate this thesis to my Grandad, who sadly passed away at the beginning of my PhD journey. Thank you for your profound influence upon my inquisitiveness and insatiable thirst for information and knowledge. Without these skills, I would not be where I am today. I miss our conversations.

ABBREVIATIONS

AEFS	Accumulated EvoFIT Stage	
-	**AEFS A**	Accumulated EvoFIT Stage A (Evolving the Face)
-	**AEFS B**	Accumulated EvoFIT Stage B (Holistic Scales)
-	**AEFS C**	Accumulated EvoFIT Stage C (Shape Tool)
-	**AEFS D**	Accumulated EvoFIT Stage D (Individual Feature Scales)
ANOVA	Analysis of Variance	
CI	Cognitive Interview	
CLT	Cognitive Load Theory	
CR	Context Reinstatement	
ECI	Enhanced Cognitive Interview	
ES	Effect Size	
H-CI	Holistic-Cognitive Interview	
HI	Holistic Interview	
LTM	Long Term Memory	
M	Mean	
MCR	Mental Context Reinstatement	
RBE	Recognition-Based Enhancements	
RES	Retrieval-Enhanced Suggestibility	
SD	Standard Deviation	
STM	Short Term Memory	
UCLan	University of Central Lancashire	
US	United States	
VOE	Verbal Overshadowing Effect	
WAH	Witness at Home	

CHAPTER 1: LITERATURE REVIEW

General introduction

According to data from the Office for National Statistics (2023), there has been a significant 11% increase in police-reported crime, following the lifting of COVID-19 restrictions in June 2021. This increase in criminal activity highlights the critical importance of ensuring that methods involved within criminal prosecutions (e.g., investigative interviewing techniques, identity parades, and facial composite constructions) are conducted with optimal efficiency, to guarantee accurate identification and apprehension of the correct offenders.

One fundamental aspect correlated to successful offender identification that emerges throughout eyewitness literature (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984; see Köhnken et al., 1999; Memon, Meissner et al., 2010 for a meta-analysis) is the technique of asking a witness to provide more information. Specifically, providing a witness with multiple attempts to recall details about the witnessed crime has been shown to enhance both the quality and quantity of their recollections (e.g., Brown et al., 2017; Fisher & Geiselman, 1992; Geiselman et al., 1984; Murre & Dros, 2015; Roediger & Butler, 2011; Wheeler & Roediger, 1992; see Köhnken et al., 1999; Memon, Meissner et al., 2010 for a meta-analysis). Although this technique has been extensively researched and is a component of one of the most effective (Memon, Meissner et al., 2010) memory facilitating techniques within investigative interviewing (the cognitive interview (CI): Fisher & Geiselman, 1992; Geiselman et al., 1984), this approach has yet to be explored in the context of composite construction procedures. This research gap might be a significant oversight, as enabling witnesses to retrieve more accurate information, particularly related to the face of the offender, could also permit them to construct a more accurate composite. This, in turn, has the potential to substantially increase the probability of correct offender identification, resulting in higher conviction rates, and reduced instances of misidentifying innocent individuals (e.g., Innocence Project, 2023).

The literature review in this thesis follows a thematic structure. It commences by delving into common themes encompassing eyewitness memory, including discussions on criminal procedures and current crime statistics. The review then transitions its focus toward composite systems, providing an in-depth description of the EvoFIT software, the system employed for constructing the composites in this thesis. Subsequently, the literature review explores specific factors that hinder (e.g., anxiety, cognitive load, and retention intervals) and facilitate (e.g., the components of the CI, and the holistic-cognitive interview (H-CI) witness memory. This comprehensive review of existing literature lays the foundation for a deeper understanding of the experiments throughout the thesis.

Innocence Project

To emphasise the sheer importance of investigating the techniques that can enhance the recollection of witnesses' memories, it is imperative to examine the compelling statistical insights provided by the Innocence Project (2023). The Innocence Project, situated in the US, is an organisation dedicated to exonerating individuals, primarily using DNA evidence, who have been wrongly convicted and are serving sentences for crimes that they did not commit. It is fundamental to remember that the application of DNA, as a form of evidence, only gained prominence in the 1980s (Panneerchelvam & Norazmi, 2003). Consequently, crimes predating this time relied exclusively on alternative evidentiary methods, predominantly eyewitness testimonies (Wells et al., 2002). Whilst DNA has revolutionised many criminal cases, it is important to recognise that not all crimes can be resolved using this method, and that the legal system still commonly relies on witness testimonies to gather information about the crime, and the physical attributes of the offender (Innocence Project, 2023; Wells & Olson, 2003; Wells et al., 2006).

In 2023 the Innocence Project documented a total of 245 successful exonerations of individuals who were wrongly convicted; alarmingly, 9% of these exonerations involved individuals who had been sentenced to death. Among these successful cases, a majority, accounting for 199 instances, were achieved through the use of new DNA evidence. Conversely, a significant discovery emerged in that 64% of these wrongful convictions were a consequence of witness misidentification. Upon further investigation, it became evident that facial composites played a role in approximately 30% of these identifications. Unfortunately, these wrongful convictions led to an additional 99 crimes committed by the true offenders, including 54 sexual assaults, 22 murders, and 23 other violent offences.

The Innocence Project provides compelling support for the need to investigate and critically evaluate current methods employed for identifying and locating offenders, predominantly concerning inaccuracies in witness memory. From the evidence presented, it is clearly illustrated that there is an urgent need for improvements in these investigative techniques (e.g., such as refining investigative interviews and composite construction procedures) to mitigate misidentifications, wrongful convictions and prevent further crimes from being committed by the true offenders. In alignment with this thesis, these findings resoundingly emphasise the imperative need to investigate eyewitness memory, investigative interviewing techniques and composite construction procedures. Despite notable developments in both techniques and technology within criminal procedures, the undeniable need for continuous improvement to prevent wrongful convictions remains indisputable.

Criminal procedures

The procedures employed to identify, locate, and ultimately convict an offender following a crime exhibit considerable variability, dependent upon the type of obtainable evidence (Frowd, 2021; Frowd, Carson, Ness, Richardson et al., 2005; Lee & Pagliaro, 2013; Wells & Hasel, 2007). When a crime occurs, the preliminary step involves searching the crime scene for any available evidence (e.g., DNA, CCTV footage, fingerprints and objects left behind). Nevertheless, there are instances where such physical evidence proves insufficient in assisting the police in the identification of the offender, or worse, there are no physical traces of evidence left behind (Frowd, Pitchford et al., 2012). In such circumstances, the police rely profoundly upon the memory of witnesses (Frowd et al., 2004; Frowd, Hancock et al., 2011; Frowd et al., 2013).

Throughout the process of gathering evidence from witnesses, the police initially conduct investigative interviews to elicit accounts of the crime, and descriptions of the offender (Chae, 2010). The predominant aim of these interviews is to gather evidence that assists in identifying and locating the offender (Wells et al., 2006). Nevertheless, an ongoing discussion persists concerning the accuracy and reliability of witness memory, as it is often asserted that human memory is both unpredictable and unreliable (Wixted, 2018; Wixted et al., 2018). This complexity presents a challenge for composite construction, as inaccurate information regarding the offender's appearance will heighten the risk of inaccuracies within the composite itself (Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Frowd, Skelton et al., 2012). Consequently, this can complicate the process of offender identification and potentially account for miscarriages of justice as demonstrated by cases highlighted within the Innocence Project (2023). However, it is important to acknowledge that the accuracy of witness memory and the accuracy of the composite construction procedures may be improved by considering the application of a diversity of factors including the use of a CI (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984), a H-CI (e.g., Frowd, Bruce, Smith et al., 2008), using multiple retrieval attempts (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984; see Köhnken et al., 1999; Memon, Meissner et al., 2010 for a meta-analysis) or shorter retention intervals (e.g., Deffenbacher, 2008; Ellis et al., 1980; Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; see Frowd et al., 2015 for a meta-analysis).

This objective stands as a central aim of this thesis, aiming to explore both investigative interviewing techniques and composite construction processes. The goal is to pinpoint stages where improvements are warranted to enhance the efficacy of witness memory, whilst reducing the probability of errors occurring.

Facial composites

When the police rely on the memory of witnesses to obtain evidence related to the identity of offenders, it typically involves recalling specific physical characteristics that pertain to the facial details of the offender (Frowd, Hancock et al., 2011; Frowd, Nelson et al., 2012; Wells & Hasel, 2007). Following the witnesses' engagement in an investigative interview to retrieve information, a composite system is then employed. Examples of such systems include E-FIT, EvoFIT, FACES and PRO-fit (see Fodarella et al., 2015 for a detailed account of the systems). These systems are used to construct a likeness of the offenders' face, known as a facial composite (Fodarella et al., 2015; Frowd et al., 2019). Importantly, these composite images are constructed solely from the memories of witnesses, with guidance from forensic practitioners (Zahradnikova et al., 2018). The construction of a composite typically takes around an hour (Frowd et al., 2019).

The application of composite images is particularly relevant in cases of serious crimes involving violence (e.g., knife crime, murder, and sexual assault) and where locating the offender is of paramount importance (Frowd & Hancock, 2008; Frowd, Bruce, Smith et al., 2008; Frowd, McQuiston-Surrett et al., 2007; Frowd, Nelson et al., 2012). The process of constructing a composite generally occurs within one to two days after the witness has observed the initial crime. These delays are often due to factors such as the lack of police availability and resources (Fodarella et al., 2015; Frowd, Pitchford et al., 2012; Frowd et al., 2019; Frowd, Skelton et al., 2012). Once the composite has been constructed, it is then distributed throughout police forces to determine whether the offender can be identified from existing databases. If this initial attempt at identification proves unsuccessful, the composite is then released to the public, through the media, in the hope that someone might recognise him or her (Davies et al., 2000; Frowd et al., 2015; Frowd et al., 2004; Frowd, Hancock et al., 2011; Frowd et al., 2019; Wells & Hasel, 2007).

Featural and holistic processing

The task of constructing a facial composite involves two unique memory retrieval processes which are known as recall and recognition (Frowd, Bruce, Smith et al., 2008). First, recall regards the process of retrieving information that has been previously stored within an individual's memory. With regards to the process of constructing a composite, recall typically involves a witness verbalising information regarding the offender's face in order to create the composite. Recall can be a challenging process as it requires accessing information from memory with minimal cues to guide retrieval (Frowd, Bruce, Smith et al., 2008). On the contrary, recognition involves identifying previously seen information. With regards to the composite construction process, recognition occurs when a witness sees a familiar face, such as that of the offender. Recognition tends to be easier than recall as prompts are available to facilitate memory retrieval, making the memories more accessible (Frowd, Bruce, Smith et al., 2008).

In relation to both recall and recognition retrieval processes, a range of systems are available for the construction of a composite (e.g., E-FIT, EvoFIT, FACES, PRO-fit; see Fodarella et al., 2015 for a detailed account of the systems), each with distinct procedures which contribute to determining the accuracy of the resulting composite (Fodarella et al., 2015; Frowd, Carson, Ness, Richardson et al., 2005). Given the stakes (i.e., misidentification) understanding the strengths and limitations of each system is of critical importance to ensure composite accuracy.

Before advances in technology, the initial approach for constructing composites involved the method of sketch (Fodarella et al., 2015; Frowd, Carson, Ness, Richardson et al., 2005). This technique required witnesses to sit down with a forensic sketch artist to construct a composite. During this process, witnesses recalled the facial details of the offender, while the forensic sketch artist translated the description into a composite image (Fodarella et al., 2015). However, this technique heavily relied on the availability of skilled, forensic sketch artists, and was also time consuming. As a result, alternative systems (e.g., Identi-kit and Photofit) were developed to eradicate the need for skilled forensic sketch artists. These systems employed arrays of facial features on cards, enabling witnesses to select features that closely resembled the offender's face, which were then arranged to form a composite (Davies et al., 2000; Fodarella et al., 2015).

Technological progress subsequently led to the introduction of computer-based systems, broadly categorised into feature-based and holistic-based approaches (Fodarella et al., 2015). Feature-based systems (e.g., E-FIT, FACES and PRO-fit) solely relied upon witnesses recalling individual facial features of the offender (Fodarella et al., 2015; Zahradnikova et al., 2018). It is important to note, however, that the process of recall is often challenging for individuals (Laughery et al., 1986; Shepherd & Ellis, 1996) and potentially leads to the construction of low-quality composites (Frowd et al., 2010). One example of the featural composite systems is PRO-fit. The system operates by focusing upon individual facial features (e.g., eyes, nose and mouth). At the beginning of the composite construction session the practitioner will administer either a CI or a H-CI (this will be addressed further throughout the thesis) to obtain a facial description of the offender from the witness to input into the software. After the physical description has been obtained, the practitioner then will go through, in detail, the procedure of constructing a composite. Once explained, the practitioner then opens the software and selects a database based on the description from the witness (i.e., age, ethnicity, and gender). Once the database has been loaded, the practitioner will input the information into the software. For each individual facial feature there are between 150-500 examples. To narrow down the examples, to a more manageable amount (e.g., 20), practitioners use the witnesses' descriptions of the features, for example short, average and long, for the nose. If the witness struggles to recall, the choice 'average' can be used. Witnesses then, with guidance from the forensic practitioner, go through all the options identifying which features represent the greatest likeness towards the offender's face (Fodarella et al., 2015; Zahradnikova et al., 2018). Witnesses are also given the opportunity to change the shape, sizing and positioning of the features on the face, as well as the contrast and brightening of the features. Once the witness has reached the optimal level of likeness for the facial features, the opportunity to add external features (e.g., hair) if needed, is provided. Once the witness has achieved the optimal level of likeness the composite is then considered complete and ready for distribution by the police.

In contrast, holistic-based systems (e.g., E-FIT V and EvoFIT, which will be discussed in the next section) take a different approach, in that they do not depend upon the recall of individual facial features (Frowd et al., 2004). Instead, these systems follow a holistic-based (or recognition-based) principle, mirroring how the human mind encodes and recognises facial details (Frowd, Bruce, Smith et al., 2008). Holistic systems allow witnesses to observe arrays of different faces, or facial features, prompting them to recognise which face resembles the greatest likeness to that of the offender (Fodarella et al., 2015). Upon reviewing both approaches, it is suggested that systems focusing on the recall of individual facial features may produce less accurate and reliable composites (Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008). In fact, EvoFIT produced composites four times more accurate than those constructed using feature-based systems (Frowd et al., 2015).

Throughout research it has been discovered that faces are processed configurally, meaning that they are perceived based on the location and arrangement of individual features on a face rather than the individual feature itself (Richler et al., 2011; Tanaka & Farah, 1993). The face inversion effect supports this holistic principle. This effect refers to when a face is rotated to an upside-down position (inverted), it becomes more challenging for individuals to identify. This effect underscores the dependence of the brain on holistic processing, as inverting a face disrupts the configural perception of facial information, which makes it difficult for individuals to recognise faces that are familiar. The process of inverting a face leads to interference with the configural processing of the location, and distances between, individual facial features, which demonstrates the importance of seeing faces in the typical upright context (Young et al., 1987).

Another principle supporting holistic processing is the composite face effect. This occurs when two halves of different faces are aligned to make one, new, whole face. In a study conducted by Young et al. (1987) this effect was studied with the faces of famous individuals. Results from the study indicated that naming of the top half of the face was more difficult when aligned with the bottom half of another famous face. It was theorised that when two halves of different faces are combined, it creates a whole new, unfamiliar face, rather than two familiar faces. However, when the faces are misaligned, the configuration of the face becomes disrupted, and results in individuals naming the faces with greater ease (Young et al., 1987). This again relays that faces are processed and recognised on a holistic basis and may explain why holistic systems produce more accurate composites than feature based systems (e.g., Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008). For these reasons, the composite software EvoFIT will be used for constructing the composites throughout the experiments within the thesis. The procedure of EvoFIT will now be discussed in length.

EvoFIT Procedure

The facial composite software EvoFIT, developed by Frowd et al. (2004), falls within the category of holistic-based systems. In contrast to traditional feature-based composite systems that focus on the recall of individual facial features (e.g., eyes, nose, and mouth), EvoFIT takes a holistic approach by focusing on the whole face (Fodarella et al., 2015; Frowd et al., 2004). By employing a holistic approach, EvoFIT permits witnesses to construct a composite without the need to recall and articulate individual facial features, a task that can be challenging for many witnesses (Laughery et al., 1986; Shepherd & Ellis, 1996). The central component of EvoFIT involves prompting witnesses with whole faces, allowing them to construct a composite image of the offender by recognising and selecting the faces that represent the greatest likeness to the offenders' face (Frowd et al., 2004; Frowd, McQuiston-Surrett et al., 2005; Frowd, Pitchford et al., 2012). The technique that is employed is centred around recognition-based enhancements (RBE) applied to the face. Essentially, the term RBE involves witnesses examining an array of whole faces, which triggers a sense of familiarity and enables them to choose the face that best resembles the offenders' face (Frowd et al., 2004).

The process of constructing a composite takes place at a police station (Martin et al., 2018). Before constructing the composite, witnesses participate in an investigative interviewing session where they are subjected to either a CI (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984) or a H-CI (e.g., Frowd, Bruce, Smith et al., 2008). These interviews aim to gather information about the offender's face (the CI) (e.g., Fisher & Geiselman 1992; Geiselman et al., 1984) or the offenders' facial details in combination with personality attributes (the H-CI) (e.g., Frowd, Bruce, Smith et al., 2008). The acquired information is then used throughout the composite construction process (Fodarella et al., 2015; Martin et al., 2018). Once the investigative interviewing session concludes, the EvoFIT software is then opened by the forensic practitioner. During this process, the relevant database is selected based upon the characteristics that were recalled by the witness (e.g., age, ethnicity, and gender). Once a database has been selected, the software presents an array of faces for the witness to evaluate its accuracy, in regard to its representation of the offenders' face. If necessary (e.g., the age does not accurately represent that of the offenders'), a new database can be selected; otherwise, the forensic practitioner proceeds with the composite construction process.

For the purpose of this thesis, the standard EvoFIT procedure was divided into three Accumulated EvoFIT Stages (**AEFS**). The term "accumulated" signifies that the opportunity for RBE accumulates (i.e., increases) progressively as the Stages advance. The initial Stage, involving the evolution of the face, is now referred to as **AEFS A**. Subsequently, the Holistic Scales will be referred to as **AEFS B**, and the Shape Tool referred to as **AEFS C**. As this thesis aims to investigate whether providing witnesses with more opportunity to make RBE to the internal facial features of a composite would enhance composite accuracy, a new Stage was developed by the researchers to offer witnesses an extra opportunity for added detail. This Stage mirrors **AEFS B**'s use of the Holistic Scales, but focuses on internal facial features rather than the whole face. This new Stage is termed **AEFS D**. The following section provides a comprehensive account of each **AEFS** in detail.

*Evolving the face (**AEFS A**)*: The witness is initially presented with a screen that contains 18 smooth, whole faces, where external features (i.e., ears and hair) are not visible (see Figure 1 for an example). The witness is asked to select three faces from this initial array that they consider represents the greatest likeness towards the face of the offender, with the main focus on the eye region. The emphasis on the eye region stems from the idea that other facial components can be adjusted at a later point within the construction process (e.g., Martin et al., 2018). Once the initial three faces have been selected, the witness is prompted to mentally visualise the offenders face for a few moments. Subsequently, a new screen with 18 new faces is presented, and the witness is required to choose three more faces that have the greatest likeness to the offender's face. After this selection, the witness is given another opportunity to mentally visualise the offenders face. On the next screen shown to the witness, six faces are presented. These six faces are the faces that have been previously selected by the witness. The witness is then tasked with selecting the singular face that they perceive to have the greatest likeness to the offenders' face.

Following this, a new screen of 18 textured faces is introduced, and the witness is asked to again pick three faces that they perceive to portray the greatest likeness towards the face of the offenders, still focusing exclusively on the eye region. After the selection, the witness is given time to visualise the offenders face. This process is then repeated with a new set of 18 faces. The chosen faces, from the preceding steps, are used to generate a new set of faces through an evolving process. The witness is presented with two separate screens, each containing 18 faces, and is instructed to select one face from each screen that they perceive to have the closest resemblance to the offender's face. Subsequently, the witness then narrows down their selection from two faces to one. This chosen face is then placed next to a scale of 1 (very poor likeness) to 10 (faces are identical) for the witness to rate the likeness of the face towards that of the offenders. This entire process is then repeated for "evolved" smooth faces and "evolved" textured faces. At the end, the witness is left with a "best face"; the face that they perceive to represent the greatest likeness to the offender's face. The witness is given the opportunity to repeat the face-evolving process or proceed to the next stage with their chosen "best face". In the context of this thesis, this entire procedure is referred to as **AEFS A**.

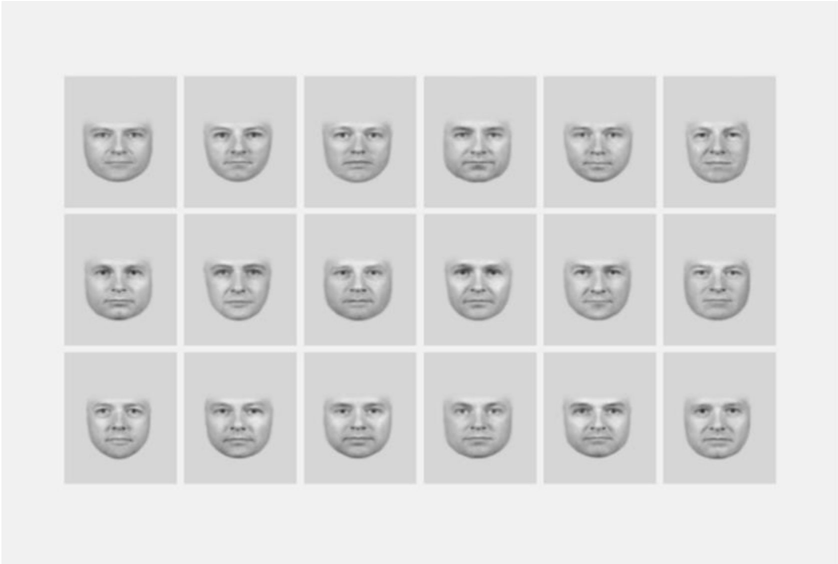
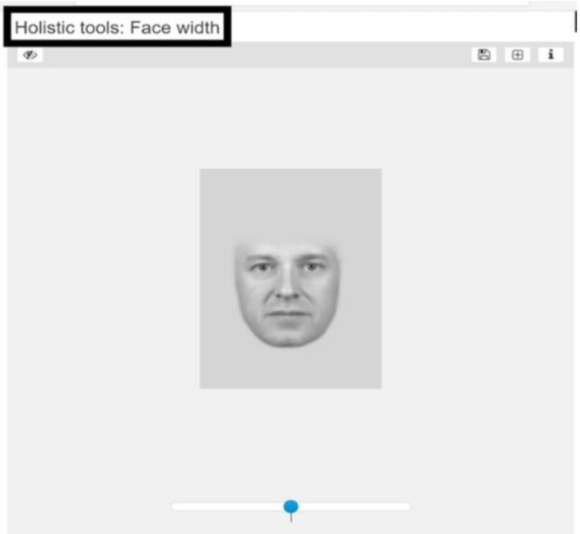


Figure 1. An example of **AEFS A**, an array of 18 composite faces.

*Holistic Scales (**AEFS B**)*: After the "best face" has been selected, the witness then proceeds to the next Stage known as the Holistic Scales. The purpose of this Stage is to make comprehensive changes in the structure of the face through a set of 15 distinct Scales.

These Scales collectively include various changes of facial appearance including face width, face weight, skin tone, age, pleasantness, outgoing, health, honesty, masculinity, threatening, trustworthiness, facial hardness, dominance, pleasantness, and face position (see Figure 2 for an example). During this process, the forensic practitioner presents each scale to the witness individually. Each scale features a slider that can be moved from left to right, which results in corresponding changes to the whole face. The witness is tasked with determining the position on each scale that best corresponds to the appearance of the offender. Alternatively, they have the option to keep the slider in the middle, indicating no modifications are necessary. Once all the changes have been exhausted, the witness's attention is directed to a more focused scale for toning adjustments. This scale, as depicted in Figure 3, provides the witness with the ability to define or undefine specific facial features (e.g., cheekbones, eyebrows, eye bags, and mouth) according to their preferences. Once all of the changes have been exhausted, witnesses have the choice to re-do the Scales, or proceed to the final of the original Stages of EvoFIT. For the purpose of this thesis, this Stage will be referred to as **AEFS B**.



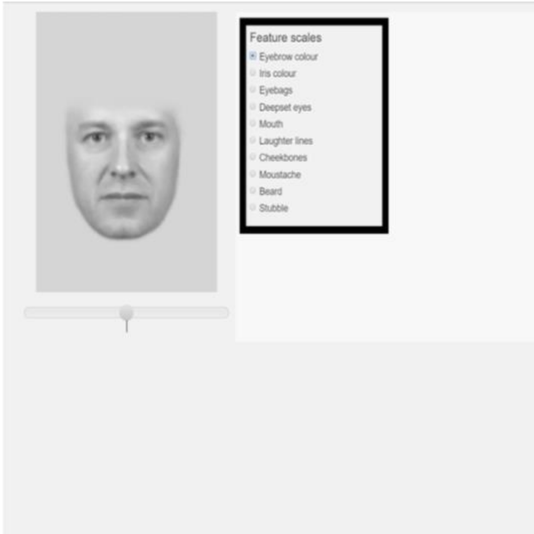
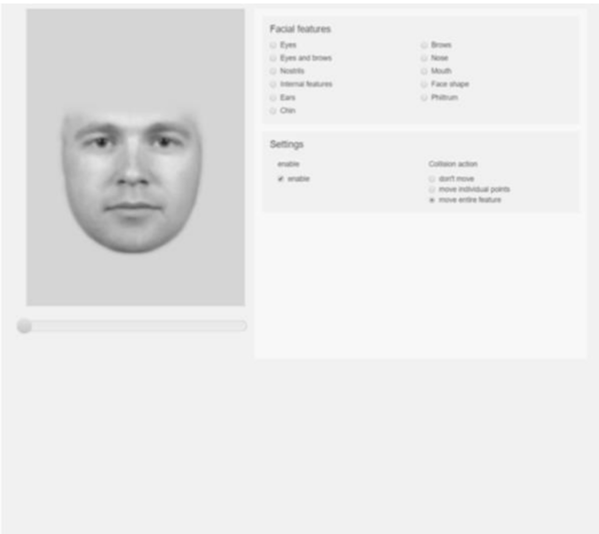


Figure 2. An example of **AEFS B**, a set of Holistic Scales for face width (left). Figure 3. An example of the Toning Scales within **AEFS B**.

*Shape Tool (**AEFS C**)*: The final of the original Stages involves the Shape Tool. The Shape Tool allows witnesses to select from an array of various features (e.g., eyes, nose, and mouth) and adjust the positioning, and sizing, of each feature on the face (see Figure 4 for an example). Once witnesses have refined the desired changes to the internal facial features, they are provided with the additional option to incorporate external features (e.g., glasses, hair, and hats) to the composite. Specifically, in regard to the hair, the witness will be presented with a small drop-down box wherein they can input relevant information regarding the hair. The database will then display a selection of hairstyle options for the witness to choose from (see Figure 5 for an example). When the witness is satisfied with both the internal facial features and the external features, in the typical EvoFIT procedure, the composite is then saved, ready for distribution. For the purpose of this thesis, the Shape Tool and external features will be referred to as **AEFS C**.

In the attempt to give witnesses more opportunity to provide RBE, an additional set of Holistic-Scales named **AEFS D** have been integrated into the software. It is well understood throughout extensive research (which will be elaborated on throughout this thesis) that providing witnesses with multiple opportunities to retrieve information about a crime, can significantly improve the quantity and quality of information that they recall (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984). Consequently, the decision was made to introduce this new Stage to the procedure to offer "witnesses" more opportunity to make RBE, ultimately leading to the construction of a more accurate composite.

*New Holistic Scales (**AEFS D**)*: The new Stage, that has specifically been developed for the purpose of this thesis is collectively referred to as **AEFS D**. Serving as an optional step, **AEFS D** grants witnesses the opportunity to use it, if they feel that they have not achieved the desired likeness from **AEFS C**. Within **AEFS D** the witness is given the opportunity to select from a wide range of individual facial features that they would like to make changes to, ranging from the typical facial features (e.g., eyes, nose, and mouth) to features such as cheekbones, chin, ears, forehead, jaw, and jowls. Once the feature had been selected for further enhancement, it undergoes the same 15 Holistic Scales from **AEFS B** (refer back to Figure 2 for an example of the Holistic Scales). Notably, unlike in **AEFS B**, witnesses are not obligated to go through all of the 15 Holistic Scales for each individual facial feature. Instead, they have the freedom to stop making RBE when they perceive they have reached the optimal likeness for the face, whether that is at the first Holistic Scale, or after exhausting all 15 of the Scales. Witnesses are at liberty to choose as many individual features as they like to make RBE to. Once witnesses have achieved the optimal likeness for the whole face the final composite will then be saved.



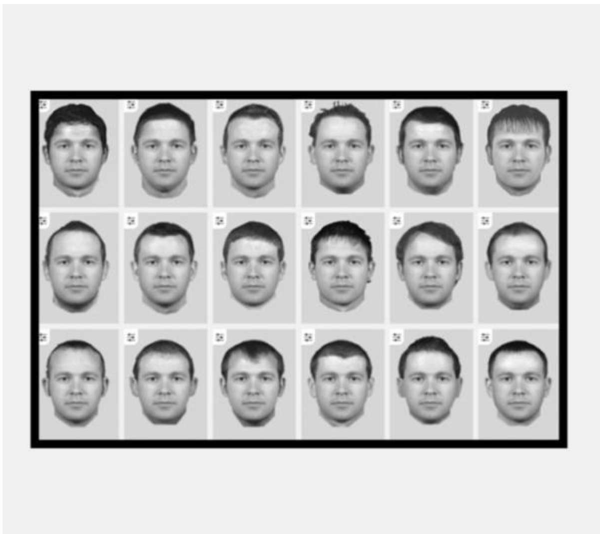


Figure 4. An example of **AEFS C**, the Shape Tool (left). Figure 5. An example of the hairstyle database (right).

Witness at Home (WAH)

The traditional face-to-face version of EvoFIT, as depicted above (e.g., Frowd et al., 2004; also see Fodarella et al., 2015 for detail), relies upon the availability of trained forensic practitioners (Martin et al., 2018). However, due to resource constraints within police forces, this approach often leads to considerable delays between a witness encoding information regarding the offenders' face, and subsequently, constructing a composite of it (Fodarella et al., 2015; Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Frowd, Nelson et al., 2012; Frowd, Pitchford et al., 2012; Frowd, Skelton et al., 2012). To overcome this issue, Martin et al. (2018) developed a more accessible version of the software called the Witness at Home (WAH) system. This adaption enables witnesses to construct a composite at the earliest convenience, and location of their choice (internet connection permitting), thus eliminating the requirement for the presence of a forensic practitioner (Martin et al., 2018). To gain access to the software, the witness is sent a unique link by a forensic practitioner. The witness will then follow the same procedure as in traditional face-to-face sessions (see the *EvoFIT* section for more detail), except they control the software themselves. The WAH procedure provides clear step-by-step instructions enabling witnesses to efficiently construct the composite without the direct guidance from a forensic practitioner. However, should the witness come across any complications with the system, they can seek assistance. As with the face-to-face approach (e.g., Frowd et al., 2019), the WAH system also takes approximately an hour to construct the composite (Martin et al., 2018). Once the witness has worked through the software and is satisfied with the likeness of the composite, they can then conclude the procedure. Completed composites are automatically saved and uploaded to the system for the forensic practitioners to review and distribute if necessary.

Despite the shared procedure for both the face-to-face and WAH systems in constructing composites, several factors impact the efficiency of these procedures and accordingly, the overall reliability of the resulting composites. One notable advantage of employing the WAH method is its convenience (Martin et al., 2018). Witnesses can construct a composite at their own, earliest convenience, without having to schedule an appointment with a forensic practitioner (Martin et al., 2018). This immediate application reduces the retention interval between the witness encoding and retrieving the offenders' facial details. The expectation is that reducing the retention interval will enhance witnesses recall of information, ultimately leading to more accurate composites (e.g., Deffenbacher et al., 2008; Ellis et al., 1980; Frowd, Bruce & Hancock, 2008; Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Frowd et al., 2015; Krouse, 1981; Murre & Dros, 2015; Shepherd et al., 1991). This influence will be discussed in more depth within the *Retention Interval* section of Chapter 1.

Furthermore, supporting the implementation of the WAH system, it is evident that the system has the potential to alleviate anxiety levels in witnesses (Martin et al., 2018). Unlike the experience of a trip to the police station, which can lead to heightened levels of anxiety and stress for a witness (Martin et al., 2017; Risan et al., 2016), the WAH system allows witnesses to construct a composite in a familiar, comfortable, environment of their choice. The WAH system removes any extra pressure that may be imposed from being in the presence of forensic practitioners within the police station, a factor which could otherwise negatively impact witness memory (Martin et al., 2018). The reduction of anxiety and stress that can be addressed through the use of the WAH system can allow witnesses to concentrate on the retrieval of their memories (Bolton & Robinson, 2017; Kieckhaefer et al., 2014; Martin et al., 2018; see Deffenbacher et al., 2004 for meta-analysis) with the potential to reduce cognitive overload. This influence will be discussed in more depth within the *Anxiety* section within Chapter 1.

However, the WAH system has limitations such as limited audience reach. The software relies on witnesses having internet access at home, which may restrict the number of individuals who can access the system. Additionally, whilst the software provides clear instructions, some witnesses may find it challenging to follow, especially if they are already feeling overwhelmed due to their involvement in a crime. The challenge of accurately navigating unfamiliar software could be seen as a difficult task and may influence the cognitive processing of witnesses when using the software (e.g., Hanway et al., 2021; Mostyn, 2012; Sweller, 1988; Sweller, 2010; Sweller et al., 2011; Sweller et al., 1990) potentially hindering memory access and compromising composite quality. This influence will be discussed in more depth within the *Cognitive Load* section within Chapter 1.

The foregoing information highlights that the accuracy of a composite is influenced by a diverse range of factors. These factors encompass variables such as: witness anxiety levels, cognitive load during interviews and composite construction, as well as the retention interval between witnessing an event and constructing a composite. Given the complexity of these factors, it is imperative to delve further into a thorough exploration of their effects. By doing so, it allows researchers to gain a deeper insight into how these factors interact with human memory and subsequently, allows practitioners to identify the most effective factors that could enhance composite procedures. These factors will now be thoroughly explored to determine their impact on witness memory and the resulting composites produced.

Factors that can influence witness memory

Anxiety: A substantial amount of research has been dedicated to investigating the connection between anxiety and the process of memory recall in witnesses (e.g., Christianson, 1992; Clifford & Scott, 1978; Dobson & Markham, 1992; Marr et al., 2021; Siegel & Loftus, 1978; Valentine & Mesout, 2009; Yuille & Cutshall, 1986; see Deffenbacher et al., 2004 for a meta-analysis). Nevertheless, despite substantial efforts, the exact impact of anxiety on witness memory remains an ongoing and evolving subject of investigation.

Researchers are particularly interested in understanding how anxiety impacts witnesses in real-world situations, where retrieval of accurate memories is of particular importance (Marr et al., 2021). This interest gains significance as crimes involving violence, which often require composites (Frowd & Hancock, 2008; Frowd, Bruce, Smith et al., 2008; Frowd, McQuiston-Surrett et al., 2007; Frowd, Nelson et al., 2012), tend to elevate levels of anxiety in both victims and witnesses (Deffenbacher et al., 2004). These heightened anxiety levels can result in challenges for witnesses with regards to the complex process of encoding, storing, and retrieving important information (Dobson & Markham, 1992). These challenges, in turn, can lead to the construction of poor-quality composites (Frowd, Bruce, McIntyre et al., 2007). Comprehending the complexities between anxiety and witness memory, therefore, does not only provide insight on the influence that it has on the cognitive processes of witnesses, but also highlights the implications for memory accuracy and composite construction more generally (e.g., Innocence Project, 2023).

There are many reasons why researchers feel that it is important to understand the impact of witness anxiety on memory retrieval, particularly in regard to criminal procedures, whereby witness testimonies play a crucial role in the identification and location of an offender (Marr et al., 2021). When witnesses are subjected to high levels of anxiety and stress while observing a crime, it can result in deleterious effects within the complex process of encoding, storing, and retrieving critical information and evidence (Bolton & Robinson, 2017; Mueller, 1979). Researchers have delved deeper into this issue (Bolton & Robinson, 2017; Marr et al., 2021; see Sauro et al., 2003 for a meta-analysis) and discovered that heightened levels of anxiety can prevent witnesses from attentively focusing on and accurately encoding information (for example, in the context of this thesis, facial details of the offender). Consequently, anxiety can disrupt the storage of the encoded information, leading to challenges during subsequent retrieval attempts, potentially leading to inaccurate details being recalled. These findings may help explain the misidentifications observed in criminal proceedings (e.g., Innocence Project, 2023).

It is important to emphasise that, whilst the impact of anxiety is of significant importance when it comes to witness memory (Deffenbacher et al., 2004), replicating levels of anxiety in studies is difficult due to ethical constraints. Nevertheless, in studies where researchers have successfully been able to replicate high levels of anxiety, noteworthy insights into the effects on witnesses have emerged. Notably, these investigations reveal that heightened levels of arousal do, subsequently, have a harmful influence on witness recognition abilities (Deffenbacher et al., 2004; Valentine & Mesout, 2009). However, according to Eysenck and

Calvo (1992), anxiety's influence on individual's task performance is recognisable only under certain circumstances. The Yerkes-Dodson law specifies that there is a U-shaped curve, which portrays the relationship between levels of arousal (e.g., anxiety and stress) and the quality of the task performed (e.g., retrieving information). According to the theoretical framework it is proposed that it is, in fact, moderate quantities of anxiety that result in the most optimum performance on a task (e.g., recalling information concerning the crime, or recognising facial details of the offender). Conversely, both heightened, and even very low levels of arousal, can have a negative impact upon witness recall (Anderson, 1994; Deffenbacher et al., 2004). This insight into the relationship between levels of anxiety and the effects that it has upon task performance is valuable for forensic practitioners, as it can assist them in utilising evidence-based techniques to reduce levels of anxiety and enhance witness memory accuracy.

Another theory that also sheds light on the relationship between anxiety and cognitive performance is the processing efficiency theory. According to this theory, different amounts of anxiety can have deleterious effects on the cognitive resources that are available when recalling information. Much like the Yerkes-Dodson law, which posits that performance is optimal with moderate levels of anxiety, the processing efficiency theory also suggests that witnesses who experience high levels of both trait and test anxiety are likely to exhibit lower levels of performance compared to those who do not. The theory stipulates that under testing conditions, anxiety is associated with poor processing, as those who are experiencing heightened levels of stress use more resources to process information than those with low levels of anxiety (Eysenck & Calvo, 1992). In regard to witnessing a crime, if witnesses experience heightened levels of anxiety whilst witnessing a crime, their attention may become directed to potential threats, rather than the crime itself or, more crucially, the facial details of the offender. Subsequently, this shift in focus could potentially hamper the encoding of information regarding the crime that has taken place, resulting in challenges during retrieval at a later point.

Timing is another vital aspect influencing memory and accuracy. In a comprehensive meta-analysis conducted by Het et al. (2005) it was emphasised that, when anxiety is experienced prior to the encoding of information (e.g., witnessing a crime and/or seeing an offenders' face) it has minimal effect on witness memory, neither improving nor impairing it. However, heightened anxiety prior to retrieval (e.g., recalling information about a crime or the identity of the offender) can adversely impact the accuracy of the information recalled. These findings are also supported by other studies (Bolton & Robinson, 2017; Shields et al., 2016) whereby it was acknowledged that heightened anxiety during retrieval can impair facial recognition at a later point.

Given this, existing research considered the implementation of evidence-based practises, such as the CI (e.g., Fisher & Geiselman, 1992) which uses the establishment of rapport to reduce witness anxiety (discussed in the *Rapport* section of this Chapter). This technique aims to establish a comfortable environment for the witnesses, thus helping them to feel more at ease throughout the investigate interview (e.g., Fisher & Geiselman, 1992; Fisher & Geiselman, 2010; Memon et al., 1995). Moreover, if anxiety during retrieval is a key factor which influences the recognition of accurate information, forensic practitioners can better manage it by implementing anxiety reducing techniques at these pivotal points (e.g., during investigative interviews, police line-ups and composite construction procedures) allowing them to exercise an amount of control over memory recall. Effectively managing witness anxiety could help to facilitate witness memory leading to the retrieval of more accurate information about the crime and better recognition of facial features.

Cognitive load: Another factor that considerably impacts the accuracy of witness memory to a crime is the cognitive load experienced during the processing of information. The term cognitive load refers to the mental strain that is placed on an individual's working memory capacity while they engage in tasks (Hanway et al., 2021; Mostyn, 2012). Working memory possesses a limited storage capacity for learning new information (e.g., crime scene information or facial details of the offender), allowing only a small amount of information to be retained, unless actively rehearsed (Eysenck & Calvo, 1992; Eysenck et al., 2007; Sweller, 2010). If information is not actively rehearsed, the amount of information that can later be successfully retrieved becomes restricted, which puts a demand on an individual's cognition, increasing the mental effort to attempt to retrieve information (Hanway et al., 2021). Sweller (2010) proposes that, although working memory does have a limited capacity in regard to storage of information, memory has an unlimited capacity when dealing with rehearsed information. Specifically, this process of rehearsal (i.e., being tested on or verbally recalling the information over a period of time) allows the transition of information from STM to long-term memory (LTM), where it remains available for later retrieval. Recognising this limitation is essential in the context of witness memory research, as failure to rehearse important crime scene details, and offender facial details, could result in information loss. Consequently, this could compromise the accuracy of investigative interviewing, potentially leading to omission of vital information, or, worse yet, the recall of inaccurate information (Engle & Kane, 2004).

Delving deeper into the concept of cognitive load and exploring methods to enhance memory retention and retrieval, the cognitive load theory (CLT) introduced by Sweller (1988) emphasises techniques that account for the limitations on cognitive processing for memory. According to this theory, an individuals' working memory can only accommodate a limited amount of information and, when this capacity exceeds demand, cognitive overload occurs, diminishing information retention and performance (Sweller, 1988; Sweller et al., 1990). There are three types of cognitive load: intrinsic load, extraneous load and germane cognitive load (Sweller, 2010). Specifically, with regard to eyewitness memory and the process of constructing a composite, intrinsic load refers to task difficulty. If a task is complex or unfamiliar to an individual, it can heighten their intrinsic cognitive load as they require more mental effort to understand and process the information being provided. Further, extraneous load refers to the cognitive load imposed by the way in which information is presented, or the instructional design of a task, meaning that badly designed tasks may increase individual's level of distraction and can increase extraneous cognitive load making learning more difficult (Sweller, 2010). For instance, simultaneously performing multiple tasks at any given time can surpass working memory capacity (e.g., Sweller et al., 2011) leading to challenges and potentially causing relevant associated information to be overlooked. Notably, in complex tasks, excessive cognitive demand can lead to reduced performance and errors (Engle & Kane, 2004).

Paivios' dual coding theory (e.g., Clark & Paivio, 1991; Paivio, 1991; Paivio & Clark, 2006) introduces two processing systems; the verbal and nonverbal systems. The verbal system handles language-related processing, whilst the nonverbal system deals with images and visual information (e.g., pictures and diagrams). If an individual relies solely upon one channel for the processing of information (i.e., solely verbal, or nonverbal information, rather than both) it can reduce the cognitive processing capacity (Baddeley, 1998; Chandler & Sweller, 1991). To optimise cognitive processing and alleviate cognitive overload, it is recommended to engage both channels (i.e., auditory/verbal and visual/pictorial; Mayer & Moreno, 2003). Baddeley and Hitch's (1974) model of working memory categorises these channels as components, including the phonological loop, central executive, and visuospatial sketchpad. The phonological loop processes auditory-verbal and visual-verbal information. Both components rely upon the central executive, which allocates cognitive resources to the appropriate processing systems (Baddeley, 2002). The concept of the dual coding theory stipulates that individuals can understand information more effectively when they utilise both verbal and nonverbal methods to encode information. By engaging in both verbal and nonverbal encoding simultaneously, individuals create multiple cognitive representations of the same information, which facilitates more accurate retrieval of information at a later point.

This theory gains particular relevance when examining its application to witnesses of criminal events. In this context, both verbal information (such as the spoken details during the crime) and visual information (such as the details of the offenders' face) can be employed. This dual approach presents an enhanced method for encoding information comprehensively, resulting in improved accessibility during subsequent retrieval. By incorporating both verbal and non-verbal techniques, this approach reduces the cognitive demand on witnesses when retaining and attempting to retrieve information. In tasks like composite construction, the integration of both verbal and non-verbal channels plays a pivotal role and may highlight the rationale behind integrating memory facilitating techniques in combination with recognition-based systems such as EvoFIT. The combined strategy allows witnesses to use the verbal channel to recall information about the crime and the offender. Simultaneously, it enables them to employ non-verbal techniques, such as recognising facial features, to construct an accurate likeness of the offenders' face (see the *Discussion* section in Chapter 2 for the effects of this in composite construction).

Retention interval: In the context of this thesis, the term "retention interval" concerns the time elapsed between a witness encoding an offender's facial details and the subsequent construction of a composite image of that face (Fodarella et al., 2015; Frowd et al., 2004). Typically, in forensic settings, this interval ranges between 24-48 hours as a result of the limited availability of police resources (Fodarella et al., 2015; Frowd, Pitchford et al., 2012; Frowd, Skelton et al., 2012). However, as the retention interval between encoding and retrieval of information increases, there is a higher probability of memory decay, potentially leading to information forgetting or even the introduction of misinformation (Krouse, 1981; Murre & Dros, 2015; Shepherd et al., 1991). Studies have shown that with extended retention intervals the accuracy of correctly named composites (i.e., identification) is typically low (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; see Frowd et al., 2015 for a meta-analysis). Nevertheless, these deficits might remain unnoticed for many weeks, possibly longer (Laughery et al., 1974). This holds significant implications for forensic practitioners conducting investigative interviews, as longer retention intervals for witnesses can substantially influence memory retrieval and the accuracy of the composite that they construct.

Differences in memory performance are notable between short and extended retention intervals (Deffenbacher et al., 2008). Memories assessed 24-48 hours after encoding are notably poorer than those evaluated immediately after encoding, though these deficits can be improved with the implementation of a shorter delay (Frowd, Carson, Ness, McQuiston-Surrett et al., 2005). Consequently, this could mean that composites constructed in a controlled environment (i.e., the laboratory), which typically adopts a short retention interval, are more likely to be accurate than those produced in a natural environment (i.e., practice) (Deffenbacher et al., 2008). This is supported through a meta-analysis conducted by Frowd et al. (2015) who confirmed that composites constructed with longer retention intervals (e.g., 24-48 hours) are less accurate when compared to those constructed with a shorter retention interval (e.g., 0-3.5 hours). Ideally, witnesses should construct a composite within a few hours of encoding the offender's face to achieve reasonably accurate results. When modern composite systems (e.g., E-FIT and PRO-fit) are used with short retention intervals of just a few hours, composite accuracy increases by approximately 15%, compared to a retention interval of 24-48 hours. These findings highlight implications that can be detrimental in practice, especially when the retention interval is typically 24-48 hours. Whilst a likeness of the offender can still be achieved after a 48-hour retention interval, only a few of these constructed composites can be accurately identified (Frowd, Carson, Ness, McQuiston-Surrett et al., 2005). The reasoning for this is that, after such a delay, the composite often reflects an overall imprint, rather than an accurate analysis of specific facial features.

While a CI, a technique known to enhance witness memory has been shown to help witnesses recall more accurate details about an offender in comparison to no CI (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984), it too, leads to recalling less overall information about a target face after 24 hours (Frowd, Carson, Ness, McQuiston-Surrett et al., 2005). Nevertheless, there are techniques that can be implemented to facilitate witness memory when there has been a lengthened delay. Frowd et al. (2013) discovered that composites which were constructed after a 24-hour retention interval produced composites with a mean correct naming score of 74% with the techniques of the H-CI, and external-features blurred, implemented. This highlights that it is possible to construct good quality composites with a delay, and that delay alone may not be responsible for poor quality composites.

In support of longer retention intervals on memory, in a meta-analysis comprised of 317 experiments (Cepeda et al., 2006), it was discovered that the retrieval of information increased in line with the retention interval. The meta-analysis summarised that the 24 hours between encoding of information and recall has been shown to lead to a facilitation in memory retrieval, compared to when the gap is less than 24 hours. However, Bell et al. (2014) stated that up until the time of their research, there had not been any studies throughout the literature which had tested an interval of between 3-24 hours (cf. 0 hours or >24 hours). To defeat this, Bell et al. (2014) conducted a study in which they asked participants to match Swahili and English words, with the retention intervals of 0 hours, 12 hours within the same day, 12 hours including overnight sleep, and 24 hours. The results indicated that both a 12-hour retention interval with sleep and a 24-hour retention interval promotes LTM retention. Again, these findings highlight that retention interval alone is not always enough to understand witness memory and that other factors, in this instance sleep, may be responsible for further deterioration of the memory deficits seen.

Again, supporting longer retention intervals and memory facilitation is the forgetting curve, proposed by Ebbinghaus (see Murre & Dros, 2015 for a recent replication). The forgetting curve informs forensic practitioners, and researchers, about the optimal timing for conducting investigative interviewing to facilitate witness memory. The principle stipulates that if witnesses do not make a clear and conscious effort to retrieve, and store, new information shortly after learning it, memory of that information will decline and be forgotten. A crucial finding regarding the forgetting curve is that the amount of information that is forgotten over a lengthy retention interval can be reduced by giving witnesses multiple opportunities to rehearse and retrieve the information. According to Ebbinghaus, these retrieval sessions should occur when recall has slowly begun to fade, but not enough to where information has been forgotten (Murre & Dros, 2015). Although the retrieval interval plays a crucial role in predicting the accuracy of witness memory, it is evident that there are also other techniques, and factors, that may play an important role.

Cognitive Interview (CI)

As identified, several factors can negatively impact both witness memory and the accuracy of the composites that they construct. However, these negative effects can be reduced by implementing memory facilitating techniques during investigative interviews (Zahradnikova et al., 2018). One of the most well-recognised techniques in this regard (Memon & Gawrylowicz, 2006) is the CI (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984); a set of mnemonics which have been extensively studied and revised throughout the years to enhance the effectiveness of memory retrieval (e.g., Geiselman et al., 1984; reviewed and revised by Fisher & Geiselman, 1992). The CI technique was developed in response to the inadequacies of standard interview approaches, which failed to effectively assist witnesses in recalling accurate crime-related memories (see Köhnken et al., 1999 for a meta-analysis). The technique consists of four stages, the mnemonics are explained in turn below.

Context reinstatement (CR): The initial stage of context reinstatement (CR) focuses on the environment in which the information about the crime was encoded. One way to achieve this is by having witnesses physically return to the original environment in which an event, or task, took place, and where the information was initially encoded (e.g., Godden & Baddeley, 1975). However, it has since been found that even when witnesses simply visualise the environment (mental context reinstatement; MCR) in which the crime took place, their memory can still be facilitated and the likelihood of them retrieving correct information increases (Fodarella et al., 2021).

In the context of investigative interviewing, forensic practitioners use CR and MCR to prompt witnesses to recall the scene of the crime, including physical and psychological factors. These factors encompass various elements such as, smells, sounds, and emotions experienced at the time that the crime took place (Fisher & Geiselman, 1992). The theoretical underpinning of this approach originates from Tulving and Thomson's (1973) encoding specificity principle, which posits that memories are encoded alongside contextual information from the physical and psychological environment. Encouraging witnesses to revisit, or mentally construct the environment, during the retrieval of information helps them to access more accurate memories associated with the contextual cues of the crime (Fodarella et al., 2021).

The effectiveness of CR has been observed for facial memory (Fodarella et al., 2021). For example, witnesses are more likely to correctly recognise an offender's face if the retrieval task takes place physically (Cutler et al., 1987; Wong & Read, 2011) or mentally (Fodarella et al., 2021) in the same environment as the environment in which the initial encoding of that face occurred. A meta-analysis by Memon, Meissner et al. (2010) found that the component of MRC alone can be as effective in facilitating witness memory, as compared to the standard interview. Furthermore, it is also worth noting that research from Milne (1997) supported this and found that CR alone can be highly beneficial in helping witnesses recall accurate information, and that it can retrieve as much accurate information from witnesses as the whole technique of the CI (Memon & Higham, 1999).

Recall everything: The second mnemonic of the CI involves encouraging witnesses to provide a comprehensive account of the information that they can remember concerning both the crime and the facial details of the offender (Geiselman et al., 1984). Witnesses are encouraged to recall all information regardless of whether they personally consider it important to the investigation (Fisher & Geiselman, 1992; Memon & Higham, 1999). By reporting as much information as possible, it is theorised that accessing other, related, memories can become possible (Memon & Gawrylowicz, 2006). In the CI, the combination of CR and report everything are considered to be the most influential mnemonics (Memon, Meissner et al., 2010).

In the field of eyewitness memory and facial descriptions, the CI has gained widespread recognition as an effective memory facilitating technique for the task of recall (Memon & Gawrylowicz, 2006). However, when it comes to the process of retrieving specific details about an offender's facial features, several difficulties arise. One primary obstacle is the way that human faces are processed by the brain, which is holistically, rather than by individual features (Wells & Hryciw, 1984). This holistic processing means that faces are encoded, stored, and retrieved as a whole, making it difficult for witnesses to recall specific facial details accurately. When considering the practical implications of using the mnemonic, it becomes evident that its implementation might inadvertently hinder witnesses' ability to recall an accurate, and detailed, description of the offenders' face. This limitation in recall accuracy can subsequently prevent the construction of an accurate composite of the offenders' face. Consequently, this could contribute to the occurrence of misidentifications when the composite is published to the general public (e.g., Innocence Project, 2023).

During investigative interviewing and the composite construction process, both recall and recognition mechanisms are employed. Recall is utilised during investigative interviews to obtain a physical description from the witness. However, recall is known to be a demanding and time-consuming task, with memories prone to rapid decay (Frowd, Bruce, Smith et al., 2008). It is evident that the recall of individual facial features, and the process of facial detail recognition (which matches how faces are encoded) are two very different cognitive processes (Wells & Hryciw, 1984) that can influence the effectiveness of each other (Frowd, Nelson et al., 2012). A concept that vividly illustrates the complexity of memory, in the context of witnesses, is the verbal overshadowing effect (VOE). The VOE refers to a phenomenon where there is a decline in the ability to recognise facial details, when they have been previously recalled and described (Frowd, Nelson et al., 2012). In practical terms, the VOE effect has noteworthy implications when it comes to the construction of composites. When witnesses are asked to describe the face of an offender, they are often required to construct a composite of the face, which involves the recognition of facial features. In essence, this transition from recall to the recognition of facial details when constructing a composite, can inadvertently introduce inaccuracies into the process of constructing a composite (see *Verbal Overshadowing Effect* section in Chapter 1 and *Introduction* section of Chapter 4 for more information in regard to composite construction).

Recall from different perspectives: The third mnemonic involves asking witnesses to retrieve information about the crime from the perspective of another witness or victim (Geiselman et al., 1984; Iriye & Jacques, 2020; Memon & Gawrylowicz, 2006). To decide which perspectives should be utilised, the forensic practitioners look at the information retrieved from witnesses during the "report everything" stage and use it as a guide as to which perspective should be used (Paulo et al., 2013). An example of change of perspective would be asking a witness, who is currently feeling anxious or stressed, to retrieve information about the crime from their perspective before the crime took place, when they reported that they were feeling more relaxed (Paulo et al., 2013). The reason why witnesses are asked to retrieve information from different perspectives is that memories are retrieved through multiple routes, and encouraging witnesses to retrieve information from different perspectives allows access to these different routes, and further, access to more information (Anderson & Pichert, 1978; Geiselman et al., 1984; Paulo et al., 2013). However, there may be issues with this stage that can lead to misperception among witnesses and potentially cause them to provide inaccurate information through speculation (Memon & Gawrylowicz, 2006; Memon & Koehnken, 1992).

Order of events: Finally, the fourth stage entails recalling events in a different order, typically in reverse order. This is based on Ebbinghaus' serial position effect theory (Murre & Dros, 2015) whereby information that is encoded first (primacy effect) or last (recency effect) is more easily retrievable than information encoded in the middle of an event (Baddeley & Hitch,1993; Schweitzer & Nuñez, 2021). This suggests that if witnesses are given additional opportunity to recall the event in reverse order, they may be inclined to recall more accurate information (Fisher & Geiselman, 1992). Further, this technique permits information, that may be overlooked, to be accessed (Paulo et al., 2013). The retrieval of information is often dictated by schemas and scripts, which are patterns of cognition or actions that individuals have developed throughout their life regarding their interpretations of a specific event (e.g., a crime scene or an offender's face: Griffiths & Milne, 2010). If information regarding a crime does not fit witnesses own personal schemas or scripts, it can cause information to be overlooked and excluded. These schemas and scripts are often stored in sequential order, so changing the order that the witness saw the crime, can overturn this problem (Paulo et al., 2013). By implementing this technique into the investigative interview, it enables the witnesses own personal schemas and scripts to be overlooked, allowing the opportunity for more, accurate and reliable information to be accessed.

Enhanced Cognitive Interview (ECI)

The enhanced cognitive interview (ECI), is an extension to the original CI, developed by Fisher and Geiselman (1992). The original CI did not address communicative issues (e.g., building rapport to decrease witness anxiety and to help transfer control of the interview to the witness) and so the ECI was developed to overcome this. The original CI is based on cognitive theories (Eisenberg, 2019) whereas the ECI is also based on the principles of social psychology, for techniques such as effective communication (Eisenberg, 2019; Fodarella et al., 2015; Memon & Higham, 1999). As a result of the new mnemonics, the investigative interview is structured, rather than being individualised to each witness, meaning that each witness is interviewed following the same procedure (Eisenberg, 2019). In a study conducted by Fisher et al. (1987) involving 16 undergraduates, the original CI was compared against their newly revised CI (ECI). The results concluded that the ECI was 45% more effective at aiding recall than the original CI. The technique of the ECI uses some of the old mnemonics from the CI, with the incorporation of new mnemonics which focus on social techniques to address communication issues. The mnemonics will be outlined below in order.

Rapport: The development from the CI to the ECI involved several important changes. One of the first important modifications was the incorporation of rapport building during the investigative interviewing process (see Chapter 2 for more detail). Rapport building is a crucial aspect of the ECI as it has been shown to produce therapeutic effects on witnesses, helping to reduce their levels of anxiety and cognitive load throughout the interview (e.g., Kieckhaefer et al., 2014; Vallano & Schreiber Compo, 2011, 2015). Due to this, it is now considered one of the most recommended approaches when interviewing witnesses (Kieckhaefer et al., 2014; Wells et al., 2006).

High levels of anxiety can negatively impact witnesses working memory, hindering their ability to access important memories (Kieckhaefer et al., 2014). By establishing rapport with the witnesses, the forensic practitioner can create a more comfortable and trusting environment, which in turn, allows the witness to feel at ease and enhances their memory retrieval during the interview process (Fisher & Geiselman, 1992).

Building rapport empowers the witness to take some control of the interview, which has been shown to be beneficial throughout investigative interviewing (Memon & Gawrylowicz, 2006; Wells et al., 2006). When rapport is successfully established, the control of the interview subtly shifts from the forensic practitioner to the witness. As a result, the witness can set the speed of the interview and recall information as they see fit (Memon et al., 1997). Open-ended questions are often used to encourage witnesses to provide a detailed and accurate account, and it is important that the forensic practitioner does not interrupt the witness during their recall. Furthermore, the questions that the witnesses are asked by the forensic practitioner should be related to the information that the witness is recalling, aligning with the witness's retrieval patterns, rather than introducing unrelated or leading questions.

Despite the importance of rapport building, studying it remains challenging for researchers due to the lack of a universally agreed-upon definition (Kieckhaefer et al., 2014; Richardson & Nash, 2022). Both verbal and non-verbal techniques can be employed during the interview process (St-Yves, 2006). Verbal techniques involve effective communication, attentive listening to the witness's statements, and the intentional use of their name, all of which contribute to creating a supportive atmosphere (Kieckhaefer et al., 2014; Vallano & Schreiber Compo, 2015). On the other-hand, non-verbal techniques, such as nodding and maintaining eye contact play a crucial role in signalling to the witness that the forensic practitioner is actively listening and engaged in the conversation (Duggan & Parrot, 2001).

Rapport building is an essential component of the ECI, as it helps witnesses to feel more at ease, improves their cooperation and enhances their ability to recall accurate information during the interview (Fisher & Geiselman, 1992), a task which most individuals find difficult. Verbal and non-verbal techniques are both valuable tools for establishing rapport, and creating a helpful environment for successful witness interviews. However, it is important to note that, besides the positive effects of the technique for investigative interviews, it has not yet been explored with regard to the composite construction process (see Chapter 3 for more information).

Explain the aims of the session: After rapport has been established, the forensic practitioner will explain the procedure to the witness, a task known to further develop rapport between the witness and forensic practitioner (Fisher & Geiselman, 1992). It will be explained that they will first take part in an interview in where they will provide the forensic practitioner with as much detail as possible about the crime, more specifically, the facial details of the offender (Fisher & Geiselman, 1992), usually followed by constructing a composite.

Context reinstatement (CR) with free recall: Once the aims of the session have been explained, CR (as used in the original CI) is used to help witnesses recall more information about the crime and the identity of the offender (refer back to the *Cognitive Reinstatement* section for more information).

Cued recall: Cued recall is where information that has been freely recalled is used at a future retrieval opportunity, specifically, as a cue, to prompt the retrieval of more information (Raaijmakers & Shiffrin, 1992). This task is carried out so that forensic practitioners can gather more information from witnesses. Cued recall regards the retrieval of information from the LTM (Memon & Higham, 1999). With regard to composite construction, this would include the forensic practitioner using information that witnesses have previously recalled as a prompt; for example, "you said the eyes were small and brown in colour. Can you remember anything else about them?" (e.g., Fodarella et al., 2015). When linking back to Tulving & Thomson's (1973) encoding specificity principle (for more information refer back to the *Context Reinstatement* section) retrieval cues are determined on whether they were present or not at the encoding of information. Cued recall can be superior to free recall (Guzel & Higham, 2013). During free recall, witnesses use their own prompts to recall information, however, if free recall is not as accurate as cued recall, it may suggest that people's own cues are not as effective (Guzel & Higham, 2013).

Repeated recall attempts: The final key component of the CI involves providing witnesses with multiple attempts to recall information (e.g., Fisher & Geiselman, 1992). Throughout a criminal procedure and investigative interviewing, it is evident that witnesses could possibly be interviewed after a number of days of the crime taking place (Memon et al., 1997) with witnesses also, possibly, being interviewed on more than one occasion (Wheeler & Roediger, 1992).

A benefit of providing witnesses with multiple retrieval opportunities is the potential that this will facilitate memory, and therefore increase the accuracy of information recalled (Chan & LaPaglia, 2011; Fisher & Geiselman, 1992; Fodarella et al., 2015; Geiselman et al., 1985; Memon, Zaragoza et al., 2010; Roediger et al., 2009; Wells et al., 2006; Wheeler & Roediger, 1992). Although giving witnesses multiple attempts at retrieval can increase the quantity of information recalled, it can also result in an increase in memory errors (Bornstein et al., 1998; Chan & LaPaglia, 2011; Henkel, 2004; McDermott, 2006; Roediger & McDermott, 1995; Roediger et al., 1996). This technique, which will be a central theme of this thesis, will be explored in greater detail throughout the next four Chapters.

The cognitive interview (CI) for composite construction

A modified version of the CI is used at the beginning of a composite construction session to help witnesses access their memories and improve the quality of the composite (Fodarella et al., 2015). For composite construction, the CI used is slightly different to the CI and ECI used in general interviewing, but is used for the same reasons: to increase the accuracy of the information in which a witness retrieves, which will further, increase the accuracy of the composites that witnesses construct (Fodarella et al., 2015).

During the composite construction interview, as with the ECI, an informal rapport is established between the witness and the forensic practitioner. Witnesses are then made aware of the process, in that it will involve an interview, in which the witnesses will provide a description of the offender's face, before constructing a composite of the face. During the recall session, witnesses are given the opportunity to visualise the offender's face and then freely recall, without interruption, unless for clarification, a description of the offender’s face. This description is recorded, by the forensic practitioner, onto a verbal-description sheet, a sheet which has headings for each individual facial feature allowing room for forensic practitioners to write down the witness's description. With this version of the CI, witnesses may also be invited to provide another recall of the facial details, however this time a cued recall task may be used, in which witnesses are prompted by forensic practitioners, with the information that they have already provided.

The development of the holistic interview (HI) and the holistic-cognitive interview (H-CI)

In the field of eyewitness memory and facial descriptions, the CI has gained widespread recognition as an effective technique for enhancing memory recall (e.g., Fisher & Geiselman, 1992). However, when dealing with the challenge of retrieving specific details about an offender’s facial features, several complexities arise. One primary obstacle lies in the holistic processing of human faces, where they are perceived as wholes, rather than by individual features (Wells & Hryciw, 1984). This holistic processing means that faces are encoded, stored, and retrieved as a whole, making it difficult for witnesses to recall specific, individual, facial details accurately.

During investigative interviews for composite construction, both recall and recognition mechanisms come into play. Recall is employed during investigative interviews to obtain a physical description of the offender, from the witness (Frowd, Bruce, Smith et al., 2008). However, it is important to note that recall can be a challenging and time-consuming task, which can result in memories deteriorating quickly (Frowd, Bruce, Smith et al., 2008). Unlike recall, which requires more effortful retrieval of information, recognition is a separate cognitive process (Baddeley, 1990) which occurs more automatically. In the context of composite construction, recognition tasks assist witnesses in identifying the composite image that most closely resembles the offenders face (Frowd, Bruce, Smith et al., 2008). Recognition is advantageous as it is a rapid and automatic process, which tends to lead to more stable and accurate memories over time (Frowd, Bruce, Smith et al., 2008). This means that witnesses are often better at recognising a face they have seen before, even if they may struggle to recall specific details about it.

Since face processing operates holistically, it is suggested that face recognition also operates holistically (Bruce & Young, 1998). Consequently, Wells and Hryciw (1984) propose that aligning the encoding and retrieving processes could enhance the accuracy of the information recalled. Successful retrieval depends on how closely retrieval cues match the cues present during encoding (e.g., holistic processing and recognition vs feature processing and recall). This indicates that a holistic approach to encoding facial features is more advantageous for recognition-based systems, which rely on identifying similarities in facial appearance. This finding reinforces the notion that holistic composite systems tend to be more accurate than feature-based systems that rely on recalling specific facial details about the offender (e.g., Fodarella et al., 2015; Frowd, Bruce & Hancock, 2008; Frowd, Bruce, Smith et al., 2008; Frowd et al., 2004; Frowd et al., 2013).

Verbal overshadowing effect (VOE): It is evident that the recall of individual facial features, and the process of facial detail recognition (which matches how faces are encoded) are two very different cognitive processes (Wells & Hryciw, 1984). It is well understood that both the recall of individual facial features, and the recognition of facial features, can influence the effectiveness of each other (Frowd, Nelson et al., 2012). VOE is an issue that arises when witnesses encode a face holistically, recall information regarding the individual facial features and then try to recognise the face at a later point (Frowd, Nelson et al., 2012). In composite construction this can lead to difficulties as witnesses typically provide a facial description of the offender's face, and then engage in a recognition-based task to select features which bare the greatest likeness to the offender's face.

Meissner and Brigham (2001) conducted a meta-analysis of 29 studies examining the effect of VOE in regard to face recognition. The meta-analysis revealed a significant, albeit small, effect of verbal overshadowing, especially when recognition was due to take place immediately after providing a description of the individual facial features. This finding supports research from Schooler and Engstler-Schooler (1990) who conducted studies showing adverse effects for recognition when witnesses were asked to recall and provide a verbal description of individual facial features. However, it has been shown that introducing a delay (e.g., as short as 24 minutes; Finger & Pezdek, 1999) can help to "release" the VOE. It was surmised that VOE may not necessarily involve a permanent impairment in memory. Instead, it is suggested that VOE might diminish the accessibility of memories. Additionally, the timing between the recollection of facial details and the recognition of the offender's face has been identified as a crucial component in this context (Finger & Pezdek, 1999). In other words, when individuals experience VOE, it is not necessarily that their memory is impaired, but rather that certain memories become harder to access within certain time frames.

Based on research findings, holistic composite systems closely emulate the cognitive processes that the human brain employs to perceives faces (Frowd et al., 2004). The integration of holistic recall, which involves capturing the overall personality of facial features, with these systems has been shown to substantially improve the accuracy of retrieved information (Wells & Hryciw, 1984). To delve deeper, the task of rating specific personality traits (e.g., aggressiveness, hostility, and intelligence) has been observed to play a significant role in increasing the effectiveness of face recognition (Berman & Cutler, 1998; Wells & Hryciw, 1984). Interestingly, the process of forming judgements about personality traits results in a higher degree of cognitive processing of facial features, ultimately resulting in more accurate details about the offenders' face, compared to the approach of solely focusing on individual facial features (Winograd, 1981). Moreover, insights from Warrington and Ackroyd (1975) suggest that making trait-based judgements about an offenders' facial features leads to greater recognition memory for faces in comparison to attempting to recall individual facial features. When translating this to composite construction, it becomes evident that making trait-based judgements of an offender's facial details may help witnesses to construct a more accurate composite of the offenders' face.

The holistic interview (HI): The holistic interview (HI) was devised in response to challenges posed by methods such as the VOE (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, Bruce & Hancock, 2008; Frowd, Bruce, Smith et al., 2008; Frowd, McQuiston-Surrett et al., 2005). The development was encouraged by research regarding the holistic processing of faces, particularly regarding the composite effect, whereby one whole face is comprised of two individual identities (i.e., one identity for the top of the face, and another identity for the bottom of the face). In a study conducted by Calder et al. (2000) composites were constructed in where the top identity displayed a different emotion to the bottom identity (e.g., anger vs happiness). Results revealed that aligning two identities to form one single composite face led to longer identification times for emotions compared to when the two identities were misaligned or inverted. This underscores the notion that emotions are processed in a configural (holistic) sense, rather than by individual features. The encoding specificity principle (Tulving & Thompson, 1974) which will be address further throughout the thesis, supports this notion. It suggests that, if the encoding (i.e., observing a face) and decoding (i.e., constructing a composite), processes share a holistic approach it can facilitate witnesses' memory. One strategy to achieve this is by prompting participants to either describe or rate personality attributes of the face as demonstrated by Berman and Cutler (1998), which laid the groundwork the development of the HI created by Frowd, Bruce, Ness et al. (2007). The HI is rooted in the theoretical underpinnings of holistic face processing, aiming to enhance memory recall within investigative environments.

The technique of the HI involved participants rating personality attributes (e.g., attractiveness, and intelligence) on a scale of low, medium, or high. Notably, one pivotal study conducted by Frowd, Bruce, Ness et al. (2007), that also laid the foundation for the subsequent development of the H-CI, involved two groups of participants. Group 1 was tasked with both describing and rating the personality traits of the target face (HI) before constructing a composite, while Group 2 were interviewed using the CI before constructing the composite. The results indicated that composites produced after the HI were approaching a significant benefit for composite accuracy, compared to those constructed with the CI. This pattern was consistent with prior research by Frowd, McQuiston-Surrett et al. (2005) in which participants observed a target face and then, 48 hours later, either engaged in no interview, a CI, or a HI before constructing a composite. The results illustrated that composites constructed after a HI were more accurate than both the CI and no interview. Notably, composites constructed from the CI were more accurate than no interview at all. This suggests that, although the rating of personality attributes helps to construct a more accurate composite than recall the facial details alone, that recall of features was still better at producing accurate composites than no intervention from practitioners at all.

Although the HI had potential benefits for composite construction (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, McQuiston-Surrett et al., 2005), it brought about a challenge for systems such as PRO-fit, which rely on the recall of facial features to locate the features on the system to construct the face of the offender (Frowd, Bruce, Smith et al., 2008). This led to the development of the H-CI which uses the principles from both the CI and HI. This adaption makes the technique more generalizable, to both composite systems that do, and do not, rely on the recall of facial details.

The holistic-cognitive interview (H-CI): The H-CI (see, Chapter 4 for more information) is a combination of both the CI and the HI (Frowd, Bruce, Smith et al., 2008). The procedure for the H-CI blends the CI techniques (e.g., rapport building, and multiple retrieval attempts of recall) with HI techniques (e.g., making personality judgements of the face). By combining the techniques, it enables the forensic practitioner to gain a physical description of the offender's face, but also uses holistic forms of retrieval, to match how the face was encoded. In the study which started the development of the H-CI, Frowd, Bruce, Smith et al. (2008) recruited two groups of participants. All participants were invited to observe a video, and then 3-4- hours later, construct a composite using the feature-based composite software PRO-fit. One group of participants engaged in a CI prior to construction, whereas the other engaged in a H-CI. Findings from the study indicated that the use of the H-CI improved the quality of both internal features (good for recognition of familiar faces; Ellis et al., 1979) and external features (good for recognition of unfamiliar faces; Young et al., 1985). The increase in accuracy for internal features is a positive find, as internal features are likely to enhance composite recognition by the police and public at a later point (Frowd, Bruce, Ness et al., 2007).

In a more recent study (Frowd, Nelson et al., 2012), the use of a holistic-based system (EvoFIT) combined with the H-CI, demonstrated that holistic judgements after a CI increased the percentage of correct composite naming, compared to the CI alone. These findings also strengthened the conclusions from Frowd, Bruce, Smith et al. (2008) who found that the H-CI was beneficial for increasing composite accuracy for both internal and external features. In regard to composite construction, the H-CI presents numerous benefits for composite accuracy, surpassing the CI in isolation.

Current research

It is evident from existing literature that the process of constructing a composite is heavily influenced by a multitude of factors that can significantly impact its accuracy. One predominant technique used within investigative interviewing is providing witnesses with multiple opportunities to recall information, thereby facilitating their memory retrieval, and potentially enhancing the accuracy of their descriptions.

Surprisingly, despite the widespread application of this technique within investigative interviewing, its potential impact on composite construction has not been thoroughly explored. This research gap serves as the primary motivation for the present thesis which aims to investigate whether providing “witnesses” with more opportunity to make RBE to the internal facial features of a composite, could lead to an increase in the overall accuracy of the resulting composite. By giving “witnesses” more opportunity to make RBE to the internal facial features during composite construction, it is proposed that it will enhance the accuracy of the resulting composite. The findings of the research may have potential benefits for forensic investigations, specifically with regards to constructing more accurate composites. This enhancement, in turn, holds the potential to increase the rates of identifying and locating the offender, while concurrently decreasing the occurrence of misidentification, a matter of considerable importance as highlighted in the Innocence Project (2023).

CHAPTER 2: EXPERIMENT 1

Asking for more information: Can recognition-based enhancements (RBE) made to internal features produce more accurate composites? *Using a Witness At Home (WAH) procedure*.

ABSTRACT

A key obstacle for composite construction research is the ability to facilitate eyewitness memory to increase composite accuracy. With crime constantly on the rise, there is a need to improve such procedures to increase successful identification of offenders and reduce misidentification rates. Research on investigative interviewing suggests that providing witnesses with multiple opportunities to retrieve information can facilitate memory retrieval (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984). The current experiment aims to test this during composite construction. In Stage 1, 30 participants were shown an unfamiliar target face (for 60 seconds). Participants were then invited back, 24 hours later, to construct a composite of the target face. Participants were assigned to one of four Accumulated EvoFIT Stage (**AEFS**) conditions, each offering different levels of opportunity to make recognition-based enhancements (RBE): **AEFS A**: involved choosing the "best" face from arrays: **AEFS B**: involved **AEFS A**, plus use of the Holistic Scales: **AEFS C**: involved **AEFS B**, plus use of the Shape Tool: **AEFS D**: involved **AEFS C**, plus use of Holistic Scales for individual internal facial features. In Stage 2, new participants were shown the composites constructed from these four **AEFS** and were asked to either name (*N* = 32) or rate their likeness to the target face (*N* = 29). In accordance with existing research, it was predicted that as the opportunity to make RBE increases (i.e., A>B>C>D) composite quality should also increase. Correct composite naming scores were unexpectedly low overall (16.6%) with no reliable difference by **AEFS**. For likeness ratings, **AEFS A** composites were reliably rated more similar to the target face. Findings did not support the hypothesis that more opportunity to make RBE would increase composite accuracy.

INTRODUCTION

Throughout criminal procedure witnesses are often called upon to provide evidence about a crime on multiple occasions (Chan & LaPaglia, 2011; Fisher et al., 2017; Henkel, 2004; Odinot et al., 2009; Shaw et al., 1995). It is widely acknowledged that this repetition of providing information can have a significant impact on the accuracy of the information recalled (Brown et al., 2017; Fisher & Geiselman, 1992; Roediger et al., 2009; Wheeler & Roediger, 1992). For instance, witnesses might be asked to recall details about an offenders' face in multiple interviews conducted during the investigation process. This practice, which involves both free recall (unprompted retrieval) and cued recall (using prompts or cues to aid memory) tasks (e.g., Fisher & Geiselman, 1992; Geiselman, 1984), has been shown to strengthen memories and improve the recall of more accurate information (Anderson, 1983; Memon, Zaragoza et al., 2010). Additionally, it can help to reduce the likelihood of misinformation and suggestibility affecting the witness's testimony (Chan & LaPaglia, 2011). However, it is important to note that some research suggests a potential downside to repeatedly seeking information from witnesses. In certain cases, asking witnesses for more details can lead to the formation of false memories (Henkel, 2004; McDermott, 2006; Roediger et al., 1996), retrieval-enhanced suggestibility (RES: Chan & LaPaglia, 2011), retrieval induced forgetting (Shaw et al., 1995) and an increase in witness confidence levels (Henkel 2004; Odinot et al., 2009).

Although the practice of providing witnesses with multiple opportunities to recall information has been extensively tested and applied in the context of investigative interviews, the approach has not been directly investigated in the context of the composite construction process. Given this gap in research this initial experiment aimed to address the contradictory findings concerning witness memory. Specifically, it investigated whether providing participants with more opportunity to make recognition-based enhancements (RBE) (see *EvoFIT Procedure* section in Chapter 1 for more detail) to the internal facial features of a composite would lead to improved accuracy in composite images.

The cognitive interview (CI)

Research has highlighted the potential advantages of offering witnesses' multiple opportunities to recall information in terms of enhancing memory and consequently improving the accuracy of the information recalled (Brown et al., 2017; Fisher & Geiselman, 1992; Geiselman et al., 1984; Roediger et al., 2009; Wheeler & Roediger, 1992). If the provision of repeated retrieval opportunities produces these suggested benefits, it is reasonable to speculate that it could positively impact memory facilitation during the process of constructing composites, an area that has not been extensively explored (Brown et al., 2017; Fodarella et al., 2015; Frowd, Bruce, Smith et al., 2008).

This perspective on offering witnesses multiple recall opportunities informed the development of the CI (Fisher & Geiselman, 1992; Geiselman et al., 1984; Wells et al., 2006). The CI is a set of mnemonic techniques employed during investigative interviewing procedures known to enhance memory retrieval (see *Cognitive Interview* section in Chapter 1 for more detail; Fodarella et al., 2015; Frowd, 2011; Memon, Meissner et al., 2010). A key component of the CI is its emphasis on repeated opportunity to retrieve information (Chan & LaPaglia, 2011) achieved by prompting witnesses to think about the crime and then recall details often multiple times during free and cued recall tasks (Fisher & Geiselman, 1992; Geiselman et al., 1984; Wells et al., 2006).

The overarching goal of the CI is to acquire more reliable and accurate evidence (Fisher & Geiselman, 1992; Geiselman et al., 1984; Wells et al., 2006). Indeed, existing literature suggests that, compared to standard investigative interviews, implementing the CI helps witnesses deliver more reliable information (Griffiths & Milne, 2010; Memon & Gawrylowicz, 2006; Milne et al., 1995; see Köhnken et al., 1999 for a meta-analysis). Milne et al. (1995) played a pivotal role in highlighting the advantages of the CI technique over the conventional standard interview (Memon & Gawrylowicz, 2006). In their research, they instructed participants to view a video and then subsequently interviewed them one week later. To examine the efficacy of the CI participants were divided into two distinct groups. Group 1 were interviewed using the CI, while Group 2 were interviewed with a standard interview. Results from the experiment revealed that Group 1 (participants subjected to the CI) exhibited a higher frequency of errors in their responses compared to Group 2 (participants subjected to the standard interview). However, this seemingly disadvantageous effect was overshadowed by an important finding: Group 1, despite making more errors, provided a substantially greater amount of accurate information compared to Group 2. In fact, the accuracies were superior to, and overshadowed, the errors (Memon & Gawrylowicz, 2006; Milne et al., 1995). This finding aligns with insights presented by Chan and LaPaglia (2011) in their review of literature, where they elucidate that the accuracy of recall in standard, unstructured interviews, tends to decline as retrieval opportunities increase. In contrast, structured investigative interviews that incorporate open-ended questions (e.g., the CI) tend to elicit more accurate information with repeated retrieval attempts.

In support of the potential benefits stemming from the CI and repeated retrieval attempts, a study conducted by Memon, Zaragoza et al. (2010) highlighted a significant advantage for witnesses who engaged in a CI before being exposed to misinformation (i.e., listening to other witnesses' accounts and reading information in the media). The findings from this study indicated that individuals who participated in a CI, before exposure to misinformation, recalled more accurate information at subsequent points compared to those who did not engage in the CI. This insight is particularly valuable for forensic investigations where witness accounts are of critical importance. However, a notable obstacle is the timing of witness interviews. In many real-life situations witnesses may only be interviewed after a number of days, increasing their risk of exposure to misinformation (e.g., Fodarella et al., 2015; Frowd, Pitchford et al., 2012; Frowd, Skelton et al., 2012). While the research regarding the CI and repeated retrieval attempts suggests it can enhance witness recall accuracy, practical challenges do emerge and need to be considered when applying these techniques in forensic settings.

Retrieval enhanced suggestibility (RES*)*

The phenomenon of multiple retrieval attempts affecting memory is a well-documented area of research. This technique, however, can also lead to what researchers term the paradoxical effect (Roediger et al., 1997). This effect involves an increase in memory errors where the information recalled may become misplaced, decay over time, or merge with other experiences and versions of events (Chan & LaPaglia, 2011; Roediger & McDermott, 1995; Roediger et al., 1996).

Building on the concept of multiple retrieval attempts, it becomes evident that the CI can reduce the effects of witness misinformation and suggestibility (Chan & LaPaglia, 2011; Memon, Zaragoza et al., 2010). Chan and LaPaglia (2011) proposed that the isolated mnemonic of repeated retrieval attempts may reveal similar findings to the full CI technique. They noted that existing research on multiple retrieval attempts typically investigated memory prior to witnesses being exposed to misinformation (e.g., Memon, Zaragoza et al., 2010). Chan and LaPaglia (2011) aimed to elaborate on this. They wanted to investigate the impact of repeated testing on witnesses after they had been exposed to misinformation to see if it influenced witness accuracy as seen with the technique of the CI (Memon, Zaragoza et al., 2010). Their research findings supported previous studies on the topic (e.g., Chan et al., 2009; Thomas et al., 2010) and revealed that repeated retrieval attempts enhanced levels of what they termed

RES. This refers to a situation where witnesses provide inaccurate information about a crime, as a way to fill in the gaps when they cannot provide accurate information (Chan et al., 2009).

It is also the case that the time interval between the initial retrieval attempt, and subsequent opportunities to retrieve information, plays a crucial role and that witnesses may incorporate misinformation into their memories of the crime during this retention interval (e.g., Chan et al., 2009). As mentioned in the foregoing Chapter with regards to composite construction where there is typically a delay of 24 to 48 hours between witnessing a crime and constructing a composite of the offenders' face (Fodarella et al., 2015; Frowd, Pitchford et al., 2012; Frowd, Skelton et al., 2012), this delay can create the opportunity for witnesses to encode misinformation. In this case, offering witnesses additional attempts to recall information during this time may reinforce the misinformation encoded during the retention interval, increasing the likelihood of the witness recalling inaccurate information. When applied to composite construction this could explain why composites are, at times, inaccurate and innocent individuals are arrested (e.g., Innocence Project, 2023).

Forgetting

The spreading activation theory, as outlined by Anderson (1983), emphasises the significance of information retrieval in establishing connections between distinct categories of information. This heightened connectivity, achieved through multiple retrieval attempts, increases the likelihood of successfully storing LTM by reinforcing the memory traces associated with it. Successful recall of information enhances the likelihood of subsequent accurate retrieval in future attempts. This concept aligns with Bjork and Bjork's (1992) proposition that retrieving information multiple times is more effective for memory enhancement than simply learning information repeatedly. The disuse theory (e.g., Bjork & Bjork, 1992) suggests that STM has a limited capacity to store new information unless actively rehearsed (Eysenck & Calvo, 1992; Eysenck et al., 2007; Sweller, 2010). Without rehearsal, the ability to remember information later on tends to decrease (Hanway et al., 2021). One explanation for this comes from retrieval induced forgetting (e.g., Murayama et al., 2014). Retrieval induced forgetting (Bjork & Bjork, 1992; Shaw et al., 1995) is a concept that occurs when the retrieval of information from one's memory impairs the ability to retrieve related, but unpractised (i.e., information not previously retrieved) information. Recalling information multiple times over a period of time can hinder the ability to recall related information that was not strengthened through the retrieval process.

Shaw et al. (1995) emphasised that, if information is not successfully recalled during the initial retrieval attempt, it may not be retrieved in subsequent attempts due to forgetting, even when witnesses are instructed to remember specific details (MacLeod, 2002). Therefore, providing witnesses with multiple opportunities to recall information is most beneficial for information that they have recalled during a previous retrieval attempt (Shaw et al., 1995). This approach aligns with research indicating that repeated opportunities for recall reinforces memories for future rehearsal, potentially preventing unrecalled information from being accessed and subsequently, forgotten (Anderson et al., 1994; Roediger & Karpicke, 2006). The timing of the initial retrieval attempt is of paramount importance. The forgetting curve (which will be discussed in more detail in the *Introduction* section in Chapter 5) addresses this issue (see Murre & Dros, 2015). Briefly, the forgetting curve illustrates that the longer the delay between encoding information and the initial retrieval attempt, the less likely witnesses are to recall accurate information. Conversely, shorter delays between encoding and retrieval enables witnesses to recall more accurate details over time. Repeated retrieval attempts serve to strengthen memories, making them more accessible at future points (Roediger & Butler, 2011; Roediger & Karpicke, 2006).

In the context of witnesses observing a crime, the first opportunity to retrieve information may take place at the crime scene; however, this is unlikely due to the availability of police resources (Fodarella et al., 2015; Frowd, Pitchford et al., 2012; Frowd, Skelton et al., 2012). If information regarding facial details is not recalled at this initial point, it is possible that it will be forgotten at future retrieval attempts (i.e., during the investigative interview or composite construction process). This may account for errors in composite construction (e.g., Innocence Project, 2023; Pitchford et al., 2017), as possible details are forgotten, resulting in inaccurate retrieval and, consequently, the production of an inaccurate composite. Therefore, it is essential to investigate methods that can increase witnesses' memory during the initial recall attempt and ensure this enhancement continues throughout subsequent retrieval attempts.

Confidence levels

The impact of repeated retrieval attempts on witness memory is a crucial aspect of memory research, as it influences the accuracy and confidence of memory recall. One study by Odinot et al. (2009) sheds light on this issue. In their research, two groups of 31 participants watched a 20-minute video and returned five weeks later. They were then asked to answer 30 open-ended questions (split into two blocks of questions; A and B, that corresponded with each other) related to the video and rate their confidence of memory recall. One group was given the opportunity to practice retrieval by answering a set of questions either one or three weeks before the final retrieval task, whilst the other group did not have a practice session. Both groups were invited back after five weeks for the official retrieval task. The findings from the study revealed that participants in the repeated retrieval condition, who had the opportunity to practice retrieval, reported higher levels of confidence in their responses. Importantly, this increase in confidence was observed for both correct and incorrect answers. This suggests that repeated retrieval attempts can lead witnesses to be more confident in the accuracy of their retrieved memories, even when they are incorrect (see also Odinot et al., 2013; Sweller, 2010).

When considering the implications of this research for cases involving composites, it becomes evident that witnesses' confidence increases for their memories even when those memories are inaccurate and this can lead to recall of misinformation and subsequently, the construction of inaccurate and ineffective composites. This aligns with statistics from the Innocence Project (2023) (see *Innocence Project* section in Chapter 1 for an in-depth review) which indicate that a substantial portion, 30% of the 375 wrongly convicted cases, involved the use of a composite. It could potentially be possible that witness's overconfidence for their retrieved memories is a contributing factor to this.

A possible explanation for this could be that increased confidence may stem from reduced anxiety levels during retrieval (Delleman & Fernandes, 2015). To illustrate this point, research from Bornstein et al. (1998) investigated the effect of violence on witness memory. Their research involved participants observing either a violent or non-violent film. After observing the film, participants were then given three attempts to retrieve information about the film. Those who were exposed to the violent film were better at providing information about the film, but were worse at recalling information around the violent event. Both sets of participants who witnessed the violent and non-violent films recalled more information with each repeated retrieval attempt, suggesting that impairments from anxiety (see *Anxiety* section in Chapter 1 for more detail) can be reduced with repeated testing. This supports research from Scrivner and Safer (1988) who also found that the deleterious effects that can be produced with anxiety after witnessing a violent crime (e.g., Bolton & Robinson, 2017; Mueller, 1979) can be reduced with multiple retrieval attempts.

In further support of the notion that repeated retrieval can reduce the effects of anxiety, Smith et al. (2016) conducted a study where participants studied new information. Participants were then split into two groups where half of the participants re-learned the information three times, and the other half of participants completed three consecutive recall tests. After a 24-hour retention interval, participants either engaged in a stress induced task, or a control task, before a final recall task. Interestingly, those participants who engaged in the re-learning information task in combination with the stress induced task, retrieved less accurate information than those in the re-learning condition in combination with a normal control task, showing that stress can impair memory. However, those participants who engaged in multiple retrieval opportunities, whether in the stress-induced task, or control task, produced similar information. This highlights the issue that multiple retrieval attempts can reduce impairments from feelings of stress and anxiety, common emotions associated with witnessing a crime (e.g., Bolton & Robinson, 2017; Frowd & Hancock, 2008; Mueller, 1979).

Current research

In various research studies (Brown et al., 2017; Chan & LaPaglia, 2011; Fisher & Geiselman, 1992; Geiselman et al., 1984; Griffiths & Milne, 2010; Odinot et al., 2006; Roediger & Butler, 2011; Roediger & Karpicke, 2006; Shaw et al., 1995; Smith et al., 2016; Wheeler & Roediger, 1992; Zaragoza et al., 2013; see Köhnken et al., 1999 for a meta-analysis), it has become evident that repeated retrieval attempts can impact both the quantity and quality of accurate information recalled by a witness. However, these techniques have only been explored in the context of investigative interviews and have not been directly explored in regard to composite construction. Given the insights from the aforementioned research, the current experiment aims to investigate whether offering "witnesses" more opportunity to make RBE (see *EvoFIT* section in Chapter 1 for more information) when constructing a composite will increase the overall accuracy of the resulting composite.

Building upon existing literature, it is hypothesised that providing participants with more opportunity to enhance the internal facial details of the composite, via RBE techniques, will result in a more accurate composite (i.e., A>B>C>D), as compared to those made with limited opportunity to do so.

METHOD

Experiment 1 comprised of two distinct Stages. The initial Stage of the experiment involved participants constructing a composite of an unfamiliar face, using a witness-led version of EvoFIT (see *EvoFIT* section in Chapter 1 for in-depth procedure). Stage 2 took place once the composites had been constructed. Stage 2 involved two, new, independent tasks (a naming task and a likeness task) that evaluated the accuracy of the constructed composites. These two Stages are explained, in detail, below.

Ethics

Experiment 1 received ethical approval from the ethics board within the School of Psychology at the University of Central Lancashire (UCLan). The ethics form was fulfilled in accordance with COVID-19 guidelines and restrictions in place at the time.

Design

The typical EvoFIT procedure was split into three Stages. Accumulated EvoFIT Stage (**AEFS) A** involved the initial section of the EvoFIT procedure, which compromised of evolving faces from arrays until a "best" face was chosen. **AEFS B** involved **AEFS A** plus the use of the 15 Holistic Scales. **AEFS C** involved **AEFS B**, plus the use of the Shape Tool. A new Stage, **AEFS D**, was created for the purpose of this research to give participants another opportunity to make RBE to the internal facial features of the composite. **AEFS D** involved **AEFS C** alongside the 15 Holistic Scales from **AEFS B**. However, these scales could be used to manipulate individual facial features, rather than the whole face (see *EvoFIT* section in Chapter 1 for the in-depth procedure). The **AEFS** are also explained in greater detail in the *Procedure* section.

Sample Size

Each Stage of the current experiment, and the following experiments, aimed to recruit ten participants per condition. The justification for this sample size is based on existing composite research (e.g., Frowd, 2021) whereby ten participants per condition has been shown to provide a strong medium effect size (Frowd, 2021).

STAGE 1: COMPOSITE CONSTRUCTION

Participants

Thirty participants, 18 females and 12 males (*M* = 25.8, *SD* = 11.0), who were unfamiliar with Coronation Street, were recruited via opportunity sampling. Due to COVID-19 social distancing restrictions, both Stages of the experiment took place online, via the share-screen option on Microsoft Teams. Participants were recruited from either the student population at UCLan using the university SONA system, a participant recruitment site whereby students earn points as part of their course credit, or from advertisements on websites: Call for Participants, LinkedIn, and Twitter. Participants from the UCLan participation system, received 7 SONA points for participation whilst other participants received a £5 Amazon voucher.

Target-photograph selection

Target-photographs were provided by Professor Charlie Frowd, from an extensive internet search. The ten (five male and five female) target-photographs were of actors and actresses from the long-standing ITV television soap Coronation Street. The identities of the characters were: Peter Barlow, Tracey Barlow, Leanne Battersby, Carla Connor, Roy Cropper, Tyrone Dobbs, Sally Metcalfe, David Platt, Kirk Sutherland and Sophie Webster. This television soap was chosen as it was thought it would be easy to reach both populations; specifically, participants who were familiar (for the naming task) and unfamiliar (for the likeness task) with the identities of the characters.

Materials and Procedure

The composite construction procedure involved two sessions that were conducted nominally 24 hours apart. During the first day of the experiment, the researcher video called the participant at their allotted timeslot. The researcher then emailed them a Microsoft Word document which contained a target-photograph (8cm x 10cm in colour) of an actor, or actress, from the television soap Coronation Street. As the experiment took place online, the document was password protected, with a complex password, so that the researcher could control when the participant accessed the document. The password was also put in place as a preventative measure, to stop the participant from re-accessing the target-photograph again, for prompts, when they were constructing the composite.

When it was time for the participant to view the composite, the researcher provided the witness with the password. As the participant needed to be unfamiliar with the target face, when the participant opened the document, they were asked to confirm if they were unfamiliar with the identity of the actor or actress. If participants were familiar with the identity, the researcher would resend a target-photograph until unfamiliarity was confirmed. The experiment adopted the WAH system for composite construction, meaning that the researcher was not involved in the process. This further meant that the researcher did not need to be blind to the target-photographs, as she could in no way interfere with construction of the composite.

Once unfamiliarity was confirmed, the researcher started a timer for the 60-second inspection period. During the 60-seconds, participants were instructed to look at the face with the aim of retaining as much detail as permissible. When the 60-seconds ended, the researcher instructed participants to close the document and put it in the recycling bin on their computer to prevent them from re-accessing the document. This was the end of the first session.

Twenty-four hours later (specifically between 20 and 28 hours) the second session took place, again via share-screen on Microsoft Teams. The researcher remained on the video call whilst the participant set up the composite software EvoFIT on their computer (this was attainable by following a unique link that was provided by the researcher) to ensure that it was running correctly. Once the software was set up, as the procedure was led by the "witness", the researcher left the video call whilst the participant constructed the composite (see *EvoFIT* section in Chapter 1 for the procedure in-depth). Although constructing the composite was led by the "witness", it is important to note that the procedure of doing so was intended to be the same as the practitioner-led version. Each screen of the software had clear instructions on what to do. The researcher was also available throughout the duration of composite construction procedure in case the participant had any questions, if he or she could not understand the instructions, or if they had any technical difficulties with the software.

Each of the 30 participants were randomly allocated to one of four **AEFS** conditions and were randomly allocated one of the ten target-photographs. Participants in **AEFS A** constructed a single composite using only the initial Stage of EvoFIT. This Stage included multiple screens, with arrays of faces. On each screen, participants were asked to select several faces, based on the eye region alone [this is standard procedure as the rest of the face can be manipulated at a later point (Martin et al., 2018)], that had the greatest likeness toward the target-photograph. The characteristics from the faces chosen were "evolved" (i.e., combined), together throughout the screens, and so as the participant progressed through the **AEFS A**, the faces they observed should start to look more like that of the target face. Participants were then instructed to choose the "best" face, the face that looked most like the target-photograph. Once the perceived "best face" had been chosen, the participant then added hair, and other external features. When the participant was satisfied that the best likeness had been achieved, the session then finished, and the system automatically saved the best face as an **AEFS A** composite. This session took approximatley 30 minutes. See Figure 6 an example of an **AEFS A** composite of Leanne Battersby.

Figure 6. An example of an **AEFS A** composite.

Participants in **AEFS B** constructed the composite in the same way as those participants did in **AEFS A**. However, for **AEFS B**, the composite also employed the 15 Holistic Scales from the EvoFIT software (Frowd, Bruce, Smith et al., 2008; Frowd et al., 2004). The Holistic Scales provided the participant with the opportunity to manipulate global aspects of the composite such as age, face weight, face width, and personality attributes such as pleasantness, health, honesty, and trustworthiness. Once the participant had worked their way through the Scales, they had the opportunity to re-do them if not satisfied with the likeness of the composite. If the participant was happy with the likeness, they then moved on to the Toning Scales. The Toning Scales provided the participant with the opportunity to manipulate the toning (greyscale colouring) of specific features (e.g., eyebrows, iris, and laughter lines) to un/define them. When the participant had exhausted any changes that they wished to make, he or she then added hair and other external features. When the participant was satisfied with the likeness of the composite, the session then finished, and the system automatically saved the face as an **AEFS B** composite. This session took approximatley 45 minutes. See Figure 7 for an example of an **AEFS B** composite of Leanne Battersby.

Figure 7. An example of an **AEFS B** composite.

The remaining ten participants were placed in **AEFS C** and **D.** The participant constructed the composite in the same way as participants in **AEFS B**. However, completing the Holistic Scales and Toning Scales, the participant was then directed to the Shape Tool. The Shape Tool allowed the participant to manipulate the sizing and positioning of the internal facial features (e.g., eyes, nose, and mouth) on the composite. The participant then added hair to the composite. The system automatically saved the face as an **AEFS C** composite. The same participant then had the opportunity to make more RBE to the face, if they felt that they had not achieved the optimum likeness of the face.

If the participant did want to make further changes to the face, they were directed to a new set of 15 Holistic Scales, which have been specifically designed and adapted by the researchers. The Scales are a replication of those from **AEFS B**. However, instead of manipulating the sizing, and personality, of the whole face, these scales permitted the participant to choose from an extensive list of individual facial features arraying from typical facial features (e.g., eyes, nose, and mouth) to features such as cheekbones, chin, ears, forehead, jaw, and jowls. The features selected are then, individually, taken through the Holistic Scales to manipulate the size, shape, positioning, and personality attributes of each feature. It is important to reiterate that these new Holistic Scales were optional, in that the participant only needed to utilise these if he or she would like to make further enhancements to the composite.

When the participant had achieved the best possible likeness of the target face, the session ended, and the system automatically saved the face as an **AEFS D** composite. This session took approximatley 60 minutes. See Figure 8 and 9 for an example of an **AEFS C** and **AEFS D** composite of Leanne Battersby.

Figure 8. An example of an **AEFS C** composite (left). Figure 9. An example of an **AEFS D** composite (right).

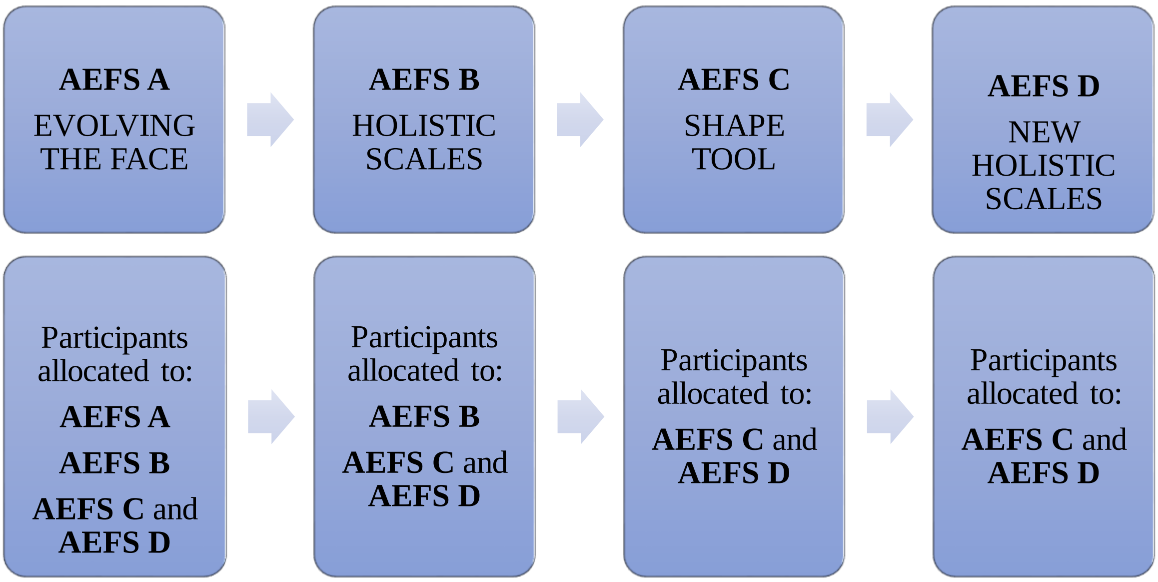


Figure 10. A figure to show participant allocation.

STAGE 2: COMPOSITE EVALUATION

NAMING TASK

The naming task involved participants looking at an array of composites and attempting to identify them.

Participants

Thirty-two participants, 26 females and six males (*M* = 41.0, *SD*= 16.0), who were familiar with the television soap Coronation Street, were recruited via opportunity sampling.

Participants were recruited in the same way as the composite construction task. Participants who were recruited from SONA received 1 SONA point for their participation, whilst other participants received a £2 Amazon voucher for their time.

Materials and Procedure

On the day of the experiment, the researcher video called the participant at their allotted time. During the video call, the researcher emailed the participant two Power-Point presentations. The first Power-Point contained the ten composites from one of the four, randomly allocated, **AEFS** conditions. The second Power-Point contained the ten target-photographs that the composites were constructed from. The second Power-Point was password protected so that the participant could not access the target-photographs before seeing the constructed composites.

Participants were requested to open the first Power-Point presentation that contained the ten composites from one of the four **AEFS**. The participant was then requested to move through the Power-Point slides, looking at each composite, saying aloud the name of the actor or actress that they thought the composite represented. If the participant was unsure, they were encouraged to guess, but could skip if they could not provide a name. The researcher made note of the answers given.

When the participant had attempted to name all of the ten composites, they were then asked to open the second Power-Point. As this was password protected, the researcher gave the participant the password. This Power-Point consisted of ten slides, and on each slide was one of the ten target-photographs used in Stage 1 of the experiment. The participant was then instructed to go through the Power-Point and name each of the target-photographs. The researcher made a note of the answers. This task was completed to ensure that the participant was satisfactorily familiar with the actors and actresses from Coronation Street. More specifically, if the participant correctly named 80% of the target-photographs their data were included within the analysis. All participants met this *a priori* rule, and so no data were excluded. There was also no evidence of missing data.

The participant was then asked to re-open the first Power-Point, to do a cued naming task to see if correct naming scores were the same, which is typically found in research, to boost power for the analysis (e.g., Frowd, Bruce, Ross et al., 2007). This task took approximatley 15 minutes.

LIKENESS TASK

The likeness task involved participants rating the accuracy of composites against the target face that they were constructed from, on a Likert Scale of 1 (low likeness) to 7 (high likeness).

Participants

Twenty-nine participants, 23 females and six males (*M* = 32.0, *SD* = 9.9), who were unfamiliar with Coronation Street, were recruited via opportunity sampling. Participants were recruited in the same way as the previous tasks. Participants who were recruited from SONA received 1 SONA point for their participation, while other participants received a £2 Amazon voucher for their time.

Materials and Procedure

On the day of the experiment, the researcher contacted participants via video call on Microsoft Teams and emailed them a Power-Point presentation. The Power-Point contained 40 slides. On each slide was one of the 40 composites (ten composites per **AEFS**) that had been constructed in Stage 1. Directly at the side of the composite was the target-photograph that the composite was constructed from. The participant's task was to rate on a Likert scale from 1 (low likeness) to 7 (high likeness) how alike the composite was to the target-photograph. The researcher made a note of the participant's answers. This task took approximatley 15 minutes.

RESULTS

ANALYSIS

The traditional Analysis of Variance (ANOVA) is adopted throughout the thesis to analyse the data. However, the traditional ANOVA does not provide the evidence needed for both the null and alternative hypothesis; therefore, Bayesian Factors ANOVA will be adopted throughout the thesis in order to provide that information.

MANIPULATION CHECK

Target-photographs were named well (*M* = 99.4%, *SD* = 2.5%) with very little difference between **AEFS** meaning that all participants had the same potential to correctly identify all composites in each **AEFS**. The naming scores are summarised in Table 1.

Table 1: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Target-Photograph Naming.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
98.8	98.8	100.0	100.0
(3.5)	(3.5)	(0.0)	(0.0)

As a manipulation check, correct target-photograph naming scores by **AEFS** were analysed using an independent samples ANOVA. There was a non-significant main effect overall on correct target-photograph naming by **AEFS** [$F(3,28) = 0.67$, $p = .580$, BF = 0.02]. This suggests that there was no significant difference between correct target-photograph naming scores by **AEFS**. The Bayes Factor (BF) of 0.02 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis.

NAMING TASK

Based upon the theoretical background discussed throughout the thesis, specifically the CI (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984) which stipulates that providing witnesses with more opportunity to provide information about a crime and/or an offender's facial details will increase the quantity and quality of information that they provide, it was hypothesised that providing participants with more opportunity to make RBE (i.e., A>B>C>D) to the facial composite would increase overall composite quality, and, further, there would be an increase in correct composite naming scores. Responses were scored for accuracy. A score of 1 was given for correct responses and 0 if no name, or an incorrect name was provided. Participants responses were checked for missing data, of which none was found. Overall, correct composite naming scores were 16.6% (*SD* = 11.2%). The naming data were analysed by participants, which analyses how many composites that each participant, individually, named within each **AEFS**. The by-participant analysis also highlights how generalisable the data is for other groups of participants (i.e., how likely another group of participants are to provide the same results). The data were also analysed by items, which analyses how often each composite was correctly named within each **AEFS.** The by-items analysis also highlights how generalisable the data is for other stimuli (i.e., if the stimuli was changed, would the results be similar in regard to the hypothesis).

By participants analysis

Correct composite naming percentages were analysed by participants. The percentage of correctly named composites, by **AEFS**, was analysed using an independent samples ANOVA. An independent samples Bayesian Factors ANOVA was also conducted to find the BF value. The naming scores are summarised in Table 2.

Table 2: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Facial Composite Naming: By Participants.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
15.0	18.8	15.0	17.5
(7.6)	(12.5)	(14.1)	(11.6)

There was a non-significant main effect on correct composite naming scores by **AEFS** [$F(3,28) = 0.21$, $p = .892$, BF = 0.01], suggesting that there was no significant difference between correct composite naming scores and **AEFS**. The BF of 0.01 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. In contrast to the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their correct naming scores.

By items analysis

Correct composite naming percentages were analysed by items. The percentage of correctly named composites, by **AEFS**, was analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The naming scores are summarised in Table 3.

Table 3: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Facial Composite Naming: By Items.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
1.2	1.5	1.2	1.4
(1.7)	(1.3)	(1.2)	(1.4)

There was a non-significant main effect on correct composite naming scores by **AEFS** [$F(3,27) = 0.10$, $p = .961$, BF= 0.01], suggesting that there was no significant difference between correct composite naming scores and **AEFS**. The BF of 0.01 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their correct naming scores.

LIKENESS TASK

As highlighted throughout the thesis, and the foregoing naming task, based on the underlying principles of the CI (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984), which highlights that the more opportunity that is given to witnesses, to provide information about a crime and/or an offender's facial details, that the quantity and quality of information that they provide will increase, it was hypothesised that providing participants with more opportunity to make RBE (i.e., A>B>C>D) to the facial composite would increase overall composite quality and, further, greater likeness ratings of the composite. Participants rated the composites from 1 (low likeness) to 7 (high likeness) against the target-photograph for likeness. The likeness data were analysed by participants, which analyses the average mean score that participants, individually, gave for the composites within the **AEFS**. The data were also analysed by items, which analyses the overall average mean score that each composite scored within each **AEFS.**

By participants analysis

Likeness scores were analysed by participants. The likeness scores of composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 4.

Table 4: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Participants.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.5	2.4	2.7	3.0
(0.8)	(0.7)	(0.8)	(0.8)

Note: The likeness scale ranged from 1 (low likeness) to 7 (high likeness). All group differences were significant, $p < .006$.

There was a significant main effect on likeness scores of composites by **AEFS** [$F(1.69,47.37) = 29.34$, $p < .001$, BF > 100]. The BF of > 100 also reveals extreme evidence (e.g., Lee & Wagenmakers, 2014) for the alternative hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did influence the accuracy of composites and likeness scores.

Follow up tests were conducted with a Bonferroni Correction ($\alpha < .05/6$). All six, paired-samples t-tests (two-tailed), revealed a significant difference between composite likeness scores and **AEFS** ($p < .006$). In more detail, in terms of the average likeness scores for each participant by **AEFS**, **AEFS A** produced significantly more accurate composites than composites constructed in **AEFS B** [with a large effect size (ES), Cohen's d = 2.1], **C** [with a large ES = 0.94] and **D** [with a medium ES = 0.68]. Furthermore, ratings became relatively worse after the "best face" was selected (**AEFS A** to **B**), then progressively improved with use of the Shape Tool (**AEFS C**) to the individual feature scales (**AEFS D**). This suggests that providing participants with more opportunity to make RBE to internal facial features did influence the accuracy of the resulting composite, although not in the direction hypothesised. It is important to note though, that although **AEFS A** produced the most accurate composites, **AEFS D** (the new additional Stage) created more accurate composites than **AEFS B** [with a large ES = .90] and, **AEFS C** (where the composite is usually saved) [with a medium ES = .65].

By items analysis

Likeness scores were analysed by items. The likeness scores of composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 5.

Table 5: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Items.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.5	2.4	2.7	3.0
(0.7)	(0.4)	(0.4)	(0.6)

Note: The likeness scale ranged from 1 (low likeness) to 7 (high likeness). One group difference was significant **AEFS A** and **AEFS B**: $p < .001$.

There was a significant main effect on likeness scores of composites by **AEFS** [$F(1.7,15.32) = 6.13$, $p = .014$, BF = 38.00]. The BF of 38.00 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the alternative hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did influence the accuracy of composites and likeness scores.

Follow up tests were conducted with a Bonferroni Correction ($\alpha < .05/6$). In more detail, regarding the average likeness scores for each composite by **AEFS**, one of the six paired-samples t-tests (two-tailed) revealed a significant difference between composite likeness scores and **AEFS** (**AEFS A** and **AEFS B**: $p < .001$ with a large ES = 1.5). The remaining five t-tests were non-significant [$p > .038$ (with an adjusted α level)]. When addressing the descriptive statistics, it is revealed that composites constructed in **AEFS A** produced more accurate composites than composites constructed in **AEFS B** (with a large ES = 1.5). This did not support the hypothesis.

DISCUSSION

The literature surrounding the subject of multiple retrieval attempts (e.g., Bornstein et al., 1998; Brown et al., 2017; Chan & LaPaglia, 2011; Fisher & Geiselman, 1992; Fisher et al., 2017; Geiselman et al., 1984; Memon, Meissner et al., 2010; Murre & Dros, 2015; Shaw et al., 1995; Smith et al., 2016) suggests that, when used correctly, this technique can enhance the accuracy of recall. However, it is important to note, as discussed earlier in this Chapter, that employing this approach can also have adverse effects. It may lead to misplacement of memories, decay, or merging of information with other experiences, ultimately resulting in inaccurate and unreliable recall (Anderson et al., 1994; Chan & LaPaglia, 2011; Chan et al., 2009; Henkel, 2004; Odinot et al., 2009; McDermott, 2006; Roediger & McDermott, 1995; Roediger et al., 1996; Shaw et al., 1995).

While this technique has been extensively investigated in the context of investigative interviewing strategies such as the CI (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984; Memon & Gawrylowicz, 2006; Wells et al., 2006; see Köhnken et al., 1999; Memon, Meissner et al., 2010 for a meta-analysis) its application to the construction of facial composites has remained relatively unexplored (cf., Brown et al., 2017). It was proposed, consistent with existing research, that providing "witnesses" with multiple opportunities to provide RBE to the internal facial features of a composite, would lead to an overall improvement in the accuracy of the resulting composites. Findings from the current experiment did not support this proposed hypothesis; instead, they revealed that composites constructed with less opportunity to make RBE, produced the most accurate composites.

Hypothesis 1, that providing participants with more opportunity to make RBE to internal facial features would increase the overall percentage of correct composite naming scores, revealed that there was no reliable difference between the accuracy of composites constructed throughout each **AEFS.** The overall percentage of correctly named composites was 16.6%.

Hypothesis 2, that providing participants with more opportunity to make RBE to internal facial features would increase the overall average likeness ratings of the composites, did not meet the hypothesis. In fact, the by participants and by items analysis revealed that composites constructed in **AEFS A**, the condition which provided participants with the least opportunity to make RBE, produced the most accurate composites, disputing the hypothesis.

However, when taking a closer look at the likeness ratings for the by participants analysis, findings revealed that **AEFS D**, the **AEFS** which provided participants with the most opportunity to make RBE, reliability produced more accurate composites than **AEFS B** and **AEFS C**, partially meeting the hypothesis. This is a potentially important finding for forensic practitioners as it suggests that, in the case of the current research, composites constructed from **AEFS C**, where the composite is usually saved, did not produce the most accurate composites.

The findings of this experiment do not align with existing literature, which typically suggests that providing witnesses with more opportunity to retrieve information leads to higher-quality and higher-quantity of information recalled about the crime (e.g., Chan & LaPaglia, 2011; Fisher & Geiselman, 1992; Fisher et al., 2017; Geiselman et al., 1984; Roediger et al., 2009; Wheeler & Roediger, 1992). Instead, the current results align with research demonstrating that providing witnesses with more opportunity to retrieve information can actually result in memory errors (Anderson, 1983; Bjork & Bjork, 1992; Chan & LaPaglia, 2011; Henkel, 2004; Köhnken et al., 1999; Memon & Gawrylowicz, 2006; Memon, Zaragoza et al., 2010; Murre & Dros, 2015; Roediger et al., 1997; Shaw et al., 1995). This unexpected outcome is particularly noteworthy when considering the success of repeated retrieval attempts in the context of investigative interviews. It does, however, bring to attention that composite accuracy does fluctuate throughout the procedure of constructing a composite, an important discovery for forensic practitioners.

There may be several reasons as to why the findings do not support the proposed hypotheses. The findings align with the theoretical concept of retrieval-induced forgetting (e.g., Shaw et al., 1995), which suggests that if information is not actively rehearsed, it is likely to be forgotten, and overridden by stronger memories, preventing more information from being recalled during each retrieval attempt. These findings further support the concept of the "forgetting curve" (e.g., Murre & Dros, 2015), indicating that memory can decay rapidly if not reinforced enough to transfer to the LTM, before forgetting sets in. As the LTM has unlimited capacity, rehearsal of information, often, will help this transfer from STM to LTM, where it stays for future retrieval attempts, aiding the retrieval process of accurate memories (Sweller, 2010).

Another explanation may regard the absence of standard interview practices (i.e., the CI: Fisher & Geiselman, 1992; Geiselman et al., 1984) before composite construction. This decision was made to isolate the impact of each **AEFS**. Combining them with memory facilitating techniques might have made it challenging to assess the isolated influence of each **AEFS**. Research from Chan and LaPaglia (2011) suggest that, if multiple retrieval attempts do not work alone to increase the accuracy of witness memory, the use of other techniques within the CI may contribute to the increase in the retrieval of accurate information. For example, the CI builds rapport to facilitate memory (e.g., Fisher & Geiselman, 1992; Kieckhaefer et al., 2014; Vallano & Schreiber Compo, 2011, 2015). As can be seen throughout comparisons of the original CI (Geiselman et al., 1984), which solely focused on cognitive theories of memory retrieval, and the ECI (Fisher & Geiselman, 1992), it is elaborated that applying social theory (e.g., improving communication between the witness and forensic practitioner) to the investigative interview can help witnesses in retrieving more accurate information (Fisher et al., 1987). Although the practitioner was involved for small parts of the procedure (e.g., during target face encoding and setting the participant up on the software), there was little opportunity to establish rapport. Literature suggests that witnesses are more likely to provide a more detailed description of an event, or identity, when they have a built rapport with the interviewer, which can only have a positive effect on composite construction (e.g., Fisher & Geiselman, 1992; Nash et al., 2016; Richardson & Nash, 2022). It may be possible that rapport is needed in accumulation with multiple retrieval attempts (e.g., the CI), to facilitate witness recall and further, the accuracy of the composite that they construct.

Another important factor to consider is the concept of divided attention and its impact on cognitive load (Eysenck & Calvo, 1992; Eysenck et al., 2007; Sweller et al., 2011). Divided attention occurs when individuals are tasked with multiple activities simultaneously, potentially overwhelming their working memory, and possibly impairing the efficiency of information retrieval (Eysenck et al., 2007). In this experiment, participants using a witness-led version of EvoFIT might have experienced an increase in their intrinsic and extraneous cognitive load due to task difficulty. When referring back to the foregoing information about Paivios' dual coding theory, it proposes that individuals process information through two channels (verbal and non-verbal) and that relying solely upon one channel for the processing of information can limit cognitive capacity (e.g., Clark & Paivio, 1991; Paivio, 1991; Paivio & Clark, 2006; Schnotz & Kürschner, 2007). In the case of the current experiment, participants were tasked with recognising facial details of the target face, navigating new software, and reading instructions simultaneously. These demands may have placed substantial overload on participants' cognitive resources for their non-verbal processing system, possibly contributing to the observed low accuracy of composites. In typical practitioner-led procedures, the forensic practitioner guides the witness through the software using verbal instructions. This guidance may reduce a witness's intrinsic and extraneous cognitive load by reducing task difficulty as they do not need to read and comprehend the written instructions simultaneously.

In future research, it would be beneficial to change the methodology of the experiment to try to further facilitate memory and thereby increase the overall accuracy of composites, so that researchers can see the potential of each **AEFS**. To achieve this aim, it would be recommended to use a practitioner-led procedure rather than a witness-led procedure. By adopting a practitioner-led methodology, it may pose many potential benefits. First and foremost, it will permit the researcher to be present throughout the whole procedure, which will hopefully help the monitoring of participants level of engagement by giving the researcher more control over the environment in which the experiment is being conducted. The practitioner being involved also has the potential to reduce any task difficulty that participants may have experienced during the witness-led procedure, by allowing participants to focus solely on the task, whilst the practitioner navigates the software and provides instructions for the participants. Further, as the researchers are now aware of the **AEFS** in isolation, without the implementation of standard interviewing procedures (e.g., the CI), using a practitioner-led method will allow the implementation of the CI, which will hopefully help to enhance composite accuracy further. By implementing the CI, it will allow the practitioner to establish rapport with the participants, a technique known to help facilitate witness memory by making participants feel at ease throughout the procedure (e.g., Fisher & Geiselman, 1992). Further, it will give the participants multiple opportunities to retrieve information about the target-face, again, another technique known to facilitate memory recall (e.g., Fisher & Geiselman, 1992).

With this in mind, Experiment 2 will adopt a practitioner-led method so that a CI can be implemented, to hopefully facilitate witness memory and therefore, the accuracy of the composite constructed. Furthermore, by involving the practitioner there is the potential to reduce task difficulty for the participants and permit them to focus on the task of constructing a composite, without navigating the software.

Experiment 1 aimed to identify whether providing participants with more opportunity to make RBE to the internal facial features of a composite increased the resulting composites accuracy. Findings did not support past research, which suggests that repeated retrieval attempts can increase the quantity and quality of information that witnesses recall. With this in mind, future research will aim to change the methodology from a witness-led procedure to practitioner-led procedure, with the addition of a CI prior to face construction. By doing so, it will hopefully help to facilitate memory retrieval and increase the accuracy of the composites more so than in the current experiment, so that the results can be taken without caution.

CHAPTER 3: EXPERIMENT 2

Asking for more information: Can recognition-based enhancements (RBE) made to internal features produce more accurate composites? *The introduction of a cognitive interview (CI).*

ABSTRACT

It is clearly demonstrated throughout an extensive body of literature that multiple retrieval opportunities, part of the CI, can increase the accuracy of witness memory (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984). The aim of the current experiment was to replicate, and extend with a CI, the findings revealed in Experiment 1. In Stage 1, 30 participants were shown an unfamiliar target-photograph (60 seconds). Twenty-four hours later, participants participated in a CI before constructing a composite of the face. Participants were allocated to one of four Accumulated EvoFIT Stage **(AEFS)**. **AEFS A**: involved selecting a "best face" from an array of faces: **AEFS B**: involved **AEFS A**, plus use of Holistic Scales: **AEFS C**: involved **AEFS B**, plus use of the Shape Tool: **AEFS D**: involved **AEFS C**, plus use of a new set of Holistic Scales for individual facial features. In Stage 2 new groups of participants were asked to either name ($N = 40$) or rate the likeness of ($N = 40$) the composites. In alignment with existing research outlined throughout the thesis, it was predicted that as the opportunity to make RBE increases (i.e., A>B>C>D) that composite quality will increase. Further, that the implementation of the CI (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984) would help to increase the percentage of correct composite naming. Correct composite naming percentages increased from Experiment 1 (27.0%) with no reliable difference by **AEFS**. For likeness ratings, **AEFS B** composites were rated more similar to the target face. Findings did not support the hypothesis.

INTRODUCTION

One of the primary challenges in investigative interviews revolves around the reliably of witness recall, a subject comprehensively discussed in prior research (Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Frowd, Nelson et al., 2012; Loftus, 2003; Wells & Hryciw, 1984; Wixted, 2018; Wixted et al., 2018). Within the topic of enhancing witness memory, a widely recognised technique (Memon & Gawrylowicz, 2006), both in research and practical applications, is the CI (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984).

Experiment 1 found that composites constructed with little opportunity to make RBE produced more accurate composites during a WAH procedure. The current experiment takes a practitioner-led approach to interviewing, with the implementation of the CI, to explore whether it, in combination with more opportunity to make RBE, would increase composite accuracy.

Rapport

The original version of the CI, developed by Geiselman et al., (1984), has been successful in improving the accuracy of information recalled by witnesses during investigative interviews (Memon & Gawrylowicz, 2006; see Köhnken et al., 1999 for a meta-analysis). However, it primarily focused on theories of cognition, and failed to address social concepts such as communicative issues between the witness and forensic practitioner. Nahouli et al. (2021) elucidate that investigative interviews are both highly cognitively and socially challenging, and that effective communication between the witness and the forensic practitioner is of upmost importance, as rapport building has been shown to reduce cognitive and social demands on witnesses during interviews. The absence of social components in the original CI led to the development of the ECI by Fisher and Geiselman (1992) which incorporated the longstanding, well established technique (Vallano & Schreiber Compo, 2015) of rapport building (Fisher & Geiselman, 1992). In the context of the current experiment, implementing a practitioner-led procedure alongside the CI, aims to minimise cognitive and social demands on participants, ultimately enhancing the accuracy of their memories, by providing a comfortable environment for witnesses, to make them feel at ease (Fisher & Geiselman, 1992; Fodarella et al., 2015; Vallano & Schreiber Compo, 2011).

It is clear that guidelines surrounding adult witness interviews recommended incorporating rapport building techniques, such as the CI, into investigative interviews to improve the quantity and quality of information provided by witnesses (Vallano & Schreiber

Compo, 2011). Rapport building is often a key factor in the success of an interview (Vallano & Schreiber Compo, 2015). The primary purpose of establishing rapport is to make witnesses feel at ease throughout the interviewing process (Fisher & Geiselman, 1992; Fodarella et al., 2015). To emphasise the benefits of rapport, a study by Collins et al. (2002) investigated its impact on the investigative interviewing procedure. Participants were assigned to one of three rapport conditions: rapport, neutral or abrupt, before recalling information (through free narrative and a semi structured questionnaire) about a mock crime that they had observed prior. The findings revealed that participants in the rapport condition recalled more accurate information. There was no reliable difference in the retrieval of accurate information between neutral and abrupt rapport conditions. These findings align with research by Vallano and Schreiber Compo (2011). Within their research, participants were tasked with observing a video of a mock crime, and then afterwards were provided with either correct or incorrect information about the crime. Participants were then exposed to different levels of rapport. Those participants who were in the rapport building condition, retrieved more accurate information about the crime, and less inaccurate information.

Although the process of building rapport in investigative interviews offers several advantages, it is not without challenges. Rapport building is known to involve both verbal and non-verbal techniques (St-Yves, 2006). Verbal techniques include creating common ground (Duke et al., 2018), demonstrating effective communication through attentiveness and verbally acknowledging witness statements (Vallano & Schreiber Compo, 2015), asking open-ended questions, and personalising the interview by using the witnesses' name (Kieckhaefer et al., 2014). Non-verbal techniques include behaviours that signal active listening through body language, such as nodding, smiling, and maintaining eye contact (Duggan & Parrott, 2001). Although there are many components known to build rapport, one major challenge arises from the absence of a globally accepted definition (Richardson & Nash, 2022). Researchers have defined rapport as a positive relationship between witnesses and forensic practitioners (Vallano & Schreiber Compo, 2015). Where the Tickle-Degnen and Rosenthal model (Tickle-Degen & Rosenthal, 1990) elaborate on this and highlight that rapport is established when there is mutual positivity, attention, and coordination between the witness and forensic practitioner. The model initially highlighted the importance of non-verbal techniques in establishing rapport, but this has since been modified to include the importance of verbal techniques, as recommended through adult guidelines (Vallano & Schreiber Compo, 2011). This new emphasis on verbal techniques is particularly relevant in the context of online data collection, due to COVID-19 guidelines, where it may be difficult to apply and interpret non-verbal techniques. The absence of a standardised procedure for forensic practitioners to establish rapport makes it challenging to know how to effectively build rapport during investigative interviews. Forensic practitioners may be unsure about which steps to implement and in what order to carry out these techniques for optimum effectiveness (Kieckhaefer et al., 2014; Vallano & Schreiber Compo, 2015). This can be problematic for composite construction procedures, as if rapport is not established, it may prevent the witness from retrieving accurate memories, resulting in the production of inaccurate composites.

The challenges associated with assessing the effectiveness of rapport building in investigative interviews are complex due to the array of factors involved. Forensic practitioners often find it challenging to determine whether they have successfully established rapport with a witness during investigative interviews (Richardson & Nash, 2022). In an effort to shed light on this issue, Richardson and Nash (2022) conducted a study evaluating the impact of rapport building within mock police interviews. In these simulated interviews, each session consisted of a mock suspect, two trainee police officers, a trainee solicitor, and an expert police interviewer, who observed through a video link. After each interview, all participants were asked to make judgments on their own personal perception of the level of rapport that had been established between the lead interviewer and the mock suspect. Interestingly, the findings revealed a consistent trend: everyone, except the lead interviewer, had a similar perception that rapport had been established at a medium to strong level. In contrast, the lead interviewer tended to disagree with the majority view and underestimated the level of rapport achieved. The authors of the study proposed that the disparity in perception between the lead interviewer and the mock suspect may be attributed to the cognitive load experienced by the lead interviewer during the process. Specifically, it is suggested that the cognitive resources required for conducting the interview may have hindered the lead interviewer's ability to accurately recognise the success of the rapport building techniques implemented.

One interesting finding from the study is that even when the lead interviewers did not personally perceive that rapport had been effectively established, the mock suspect still acknowledged presence of rapport (Richardson & Nash, 2022). This observation has potential benefits for the composite construction process, suggesting that even if forensic practitioners struggle to ascertain whether rapport has been established with a witness, as long as the witness themselves perceive it there may still be beneficial effects on memory retrieval.

However, despite its significance in investigative interviews (Fisher & Geiselman, 1992; Kieckhaefer et al., 2014; Köhnken et al., 1999; Vallano & Schreiber Compo, 2011, 2015; Wells et al., 2006), the impact of rapport, in the context of composite construction, has received limited attention. In theory, if rapport facilitates accurate memory retrieval in witnesses through its anxiety reducing effects, it has the potential to help witnesses construct more accurate composites.

Anxiety

Another way that rapport has been suggested to aid the interview process is by reducing witnesses' levels of anxiety (Kieckhaefer et al., 2014). This reduction in anxiety has been shown to enhance both the quantity and quality of information that witnesses can recall about a crime, by promoting a sense of calm and relaxation, permitting witnesses to effectively access their memory (Carter et al., 1996; Eysenck & Calvo, 1992).

Anxiety is known to hinder memory recall by disrupting working memory, a concept explained by the processing efficiency theory (Eysenck, 1979; Eysenck & Calvo, 1992). The theory elaborates that anxiety has damaging consequences on the retrieval of witness memories as it causes a depletion within the capacity of cognitive resources available, in turn, making it difficult for witnesses to conduct a comprehensive search of their memory. When witnesses cannot access their memories accurately, it results in decreased overall memory accuracy (Eysenck, 1979; Eysenck et al., 2007). In the context of this thesis, inaccurate memory recall, as a result of anxiety, may result in inaccuracies within the composite construction process, making it challenging for the public to identify the offender at a later point (e.g., Innocence Project, 2023). In the context of the current research, establishing rapport with the "witnesses" will hopefully put them at ease (e.g., Fisher & Geiselman, 1992; Fodarella et al., 2015) ensuring that multiple cognitive resources are available for accurate retrieval about the offender's face (e.g., Eysenck & Calvo, 1992). This, in turn, should enhance the accuracy of the composite they construct.

An essential consideration in rapport building is timing. Kieckhaefer et al. (2014) propose that establishing high levels of rapport before witnesses are exposed to misinformation, tends to produce more accurate information compared to witnesses who established low levels of rapport. Although high levels of rapport can reduce anxiety, which aids memory retrieval, it was suggested that the improved accuracy of information recall may not be solely due to anxiety reduction but could be linked to how rapport influences the cognitive process of memory retrieval (Kieckhaefer et al., 2014) a factor which will be discussed now.

Cognitive load

Another way in which rapport has been recognised to facilitate witness memory is by reducing their levels of cognitive load (Carter et al., 1996; Eysenck & Calvo, 1992). It is evident that a multitude of factors can overload working memory (Eysenck & Calvo, 1992; Eysenck et al., 2007). This overload is universally referred to as cognitive load, which denotes to the capacity of working memory properties that are available when an individual is engaging in any given task (Paas et al., 2003; Sweller et al., 2011). The CLT (e.g., Sweller, 1988) demonstrates how individuals efficiently complete tasks by managing the relationship between both working memory and LTM. According to this theory, individuals have two processing channels: auditory/verbal and visual/pictorial. This idea aligns with Paivio's (1986) dual coding theory and Baddeley's (1998) theory of working memory. It is essential to engage both channels when processing information, as relying solely on one can limit cognitive processing capacity (Baddeley, 1998; Chandler & Sweller, 1991). In other words, information processing benefits from using both channels (i.e., auditory/verbal and visual/pictorial) not just one (Mayer, 2002). If there is a greater, competing, mental demand placed upon witnesses' cognitive resources, it is evident that it can reduce available resources, resulting in cognitive deficits (Sweller, 1988; Sweller et al., 1990; Vredeveldt et al., 2011).

Rapport has been suggested to make witnesses feel more comfortable throughout the investigate interview (Carter et al., 1996) and therefore, decrease their level of cognitive load, by allowing more resources to be allocated to retrieving accurate information (Kieckhaefer et al., 2014). This has the potential to lead to improvements in the detection of false information and prevent witnesses from recalling it (Kieckhaefer et al., 2014), and increase the retrieval of accurate information. When there is competition for limited cognitive resources, witness performance may decline (Kieckhaefer et al., 2014). Increased cognitive load can result in decreased performance across tasks (Sweller, 2010), including retrieving accurate information related to the original memory, which was surmised in Experiment 1 (see *Discussion* section in chapter 2 for more information). To mitigate these effects, implementing the CI in the current experiment has the potential to reduce cognitive load and improve witness performance.

Current research

The aim of the current experiment is to replicate and extend the findings revealed in Experiment 1. This will be achieved by changing the methodology of the experiment, transitioning from a witness-led approach to a practitioner-led one. The involvement of a practitioner holds the potential to alleviate any task difficulties that participants may have encountered in the witness-led procedure. This transition allows participants to concentrate solely on the task, while the practitioner manages software navigate and provides clear instructions for the participants, rather than them reading the instructions themselves. Furthermore, having gained insight into the **AEFS** in isolation, without the integration of standard interviewing procedures (e.g., the CI), using a practitioner-led method will allow the implementation of the CI, which will hopefully help to enhance composite accuracy further. By implementing the CI, practitioners can establish rapport with the participants, a technique known to help facilitate witness memory by promoting participants ease throughout the procedure (e.g., Fisher & Geiselman, 1992). Moreover, the CI provides participants with multiple opportunities to retrieve information about the target-face, again, another technique known to facilitate memory recall (e.g., Fisher & Geiselman, 1992). Therefore, the practitioner-led method not only addresses potential task difficulties, but also has the potential to enhance composite accuracy by integrating the CI and promoting rapport building and multiple retrieval opportunities.

In summary, the current experiment aims to replicate Experiment 1 with the addition of a CI designed to increase composite accuracy. To achieve this, Experiment 2 will adopt a practitioner-led version of EvoFIT in conjunction with a CI. It is predicted that using these standardised methods of composite construction, in addition to providing participants with more opportunity to make RBE, will result in higher composite accuracy and more dependable outcomes compared to Experiment 1.

METHOD

As with Experiment 1, Experiment 2 was divided into two distinct Stages. Stage 1 involved the procedure of constructing a composite of an unfamiliar face, this time using a practitioner-led version of the composite software EvoFIT. Stage 2 was split into two tasks which evaluated the accuracy of the composites that were constructed in Stage 1. Task 1 involved a naming task wherein participants were asked to attempt to identify the constructed composites (from Stage 1) identities. Task 2 involved a likeness task, in which participants were asked to rate the likeness of the composite images to the target face. Stage 2 was implemented as a measure of composite accuracy.

Ethics

Ethical approval was received from the ethics board within the School of Psychology at UCLan. The ethics form was fulfilled in accordance with COVID-19 guidelines that were in place at the time of data collection.

Design

The design for Experiment 2 is the same as that in Experiment 1 (see *Design* section in Chapter 2 for more detail).

STAGE 1: COMPOSITE CONSTRUCTION

Participants

Thirty participants, 18 females and 12 males (M = 30.3, SD = 12.2), who were unfamiliar with the television soap Coronation Street, were recruited via opportunity sampling in the same way as Experiment 1. Participants recruited through the SONA website received 7 SONA points for their participation, where those recruited via external websites received a £5 Amazon voucher for their time.

Materials

The target-photographs were selected in the same way as Experiment 1 and the identities were the same as those in Experiment 1: Peter Barlow, Tracey Barlow, Leanne Battersby, Carla Connor, Roy Cropper, Tyrone Dobbs, Sally Metcalfe, David Platt, Kirk Sutherland, and Rosie Webster. The target-photographs were 8cm x 10cm and in colour.

Procedure

As with Experiment 1, Stage 1 of the current experiment took place over two consecutive days. On day one of the experiment, the researcher video called the participant, via Microsoft Teams, at their allotted timeslot. During the video call the researcher emailed participants a Microsoft Word document which pictured one of the ten target-photographs. Again, as the experiment took place over video call, the document was password protected with a complex password, so that the researcher could control when the participant accessed the document, as well as to prevent them from re-accessing the face when constructing the composite.

The researcher asked the participant to access the document and to confirm that they were unfamiliar with the identity. If familiar, the researcher would resend a target-photograph until unfamiliarity was confirmed. As the current experiment was practitioner-led it actively involved the researcher, so it was vital that the researcher was blind to the target-photographs. To achieve this, the documents were coded, so the researcher was aware of which target-photographs had been distributed.

Once unfamiliarity was confirmed, a 60-second inspection period took place where the participant was instructed to look at the target face, encoding as much detail as possible. Once the 60-seconds had ended the participant was instructed to put the document, containing the target-photograph, in the recycling bin on their computer to prevent future access. The session then came to an end.

Twenty-four hours later, the second videocall took place. First in adherence with the CI (e.g., Fisher & Geiselman, 1992), an informal rapport was established, to help ease the participant into the procedure and attempt to alleviate any possible level of anxiety that they may have. This technique has consistently been shown to facilitate recall and memory retrieval (Fisher & Geiselman, 1992; Fodarella et al., 2015; Kieckhaefer et al., 2014; Wells et al., 2006). Rapport building was achieved by the researcher asking the participant some broad questions (e.g., Duggan & Parrott, 2001; Vallano & Schreiber Compo, 2015) such as "Do you know anything about the composite construction procedure?". The researcher then thoroughly explained the procedure and informed the participant that the procedure comprised of two parts. First, that they will be asked to provide a description of the target face and second, that they will be invited to construct a composite of the face. The researcher also gave the participant the opportunity to ask any questions or request more information if needed.

The participant was invited to think about the target face using context reinstatement (e.g., Fisher & Geiselman, 1992; Fodarella et al., 2021; Tulving & Thomson, 1973; see Memon, Meissner et al., 2010 for a meta-analysis) and then they were asked to freely recall (retrieval opportunity), without interruption, as many details that they could regarding the individual facial features of the target face (e.g., eyebrows, eyes, nose, mouth, face shape). The researcher made notes of the details recalled on a verbal description sheet (e.g., Frowd, 2011) to keep a note of the description and, for if a cued recall task was to be implemented, so that the researcher could use it for cues. The CI component of the experiment took around 15 minutes to complete.

The researcher then guided the participant through the EvoFIT software. Contrary to Experiment 1, in Experiment 2 the researcher controlled the software, via screen sharing on Microsoft Teams, and provided the participant with instructions on what to do (see both *EvoFIT* section in Chapter 1, and *Procedure* section in Chapter 2 for in-depth detail).

STAGE 2: COMPOSITE EVALUATION

NAMING TASK

The naming task involved participants looking at an array of composites and attempting to identify them.

Participants

Forty participants, 29 females and 11 males (*M* = 33.0, *SD* = 15.9) were recruited via opportunity sampling. Participants were recruited in the same way as the previous Stage. Participants who were recruited from SONA received 1 SONA point for participation, whilst other participants received a £2 Amazon voucher for their time.

Materials and Procedure

The materials and procedure for the naming task were the same as in Experiment 1 (see *Naming Task* in Chapter 2 for more detail). As with Experiment 1, to ensure that participants were satisfactorily familiar with identities of the target-photographs, an *a priori* rule was specified. During the naming task, participants were shown the ten target-photographs and were asked to name them. If participants correctly named 80% of the target-photographs, their data were included within the analysis. All participants met the *a priori* specification, and so all data were included. There was also no missing data.

LIKENESS TASK

The likeness task involved participants rating the accuracy of composites against the target face on a Likert scale of 1 (low likeness) to 7 (high likeness).

Participants

Forty participants, 31 females and nine males (*M* = 27.0, *SD* = 11.7) were recruited via opportunity sampling. Participants were recruited in the same way as the previous Stages. Participants recruited from SONA received 1 SONA point for participation, whilst other participants received a £2 Amazon voucher for their time.

Materials and Procedure

The materials and procedure for the likeness task were the same as in Experiment 1 (see *Likeness Task* in Chapter 2 for more detail).

RESULTS

ANALYSIS

The traditional ANOVA is adopted throughout the thesis to analyse the data. However, the traditional ANOVA does not provide the evidence needed for both the null and alternative hypothesis; therefore, Bayesian Factors ANOVA will be adopted throughout the thesis in order to provide that information.

MANIPULATION CHECK

Target-photographs were named well (M = 99.5%, SD = 3.2%) with very little difference between **AEFS** meaning that all participants had the same potential to correctly identify all composites in each **AEFS**. The naming scores are summarised in Table 6.

Table 6: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Target-Photograph Naming.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
100.0 (0.0)	100.0 (0.0)	100.0 (0.0)	98.0 (6.3)

As a manipulation check, correct target-photograph naming scores by **AEFS** were analysed using an independent samples ANOVA. There was a non-significant main effect overall on correct target-photograph naming scores by **AEFS** [$F(3,36) = 1.00$, $p = .404$, BF = 0.02]. The BF of 0.02 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. This suggests that there was no significant difference between correct target-photograph naming scores by **AEFS**.

NAMING TASK

Responses were scored for accuracy. A score of 1 was given for correct responses and 0 if no name, or an incorrect name was provided. Participants responses were checked for missing data, of which non was found. Overall, correct composite naming scores were 27.0% (*SD* = 17.6%). The naming data were analysed both by participants and by items (see *Naming Task* section in Chapter 2 for description). It was hypothesised, in agreement with existing research (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984) that providing participants with more opportunity to make RBE (i.e., A>B>C>D) to the facial composite would increase overall composite quality, and, further, the greater the percentage of correct composite naming scores would be.

By participants analysis

Correct composite naming percentages were analysed by participants. The percentage of correctly named composites, by **AEFS**, was analysed using an independent samples ANOVA. An independent samples Bayesian Factors ANOVA was also conducted to find the BF value. The naming scores are summarised in Table 7.

Table 7: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Composite Naming: By Participants.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
29.0 (17.3)	28.0 (17.5)	28.0 (15.5)	23.0 (21.6)

There was a non-significant main effect on correct composite naming scores by **AEFS** [$F(3,36) = 0.22$, $p = .880$, BF = 0.01], suggesting that there was no significant difference between correct composite naming scores and **AEFS**. The BF of 0.01 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their correct naming scores.

By items analysis

Correct composite naming percentages were analysed by items. The percentage of correctly named composites, by **AEFS**, was analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The naming scores are summarised in Table 8.

Table 8: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Composite Naming: By Items.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
2.9	2.8	2.8	2.3
(2.8)	(1.9)	(2.3)	(1.4)

There was a non-significant main effect on correct composite naming scores by **AEFS** [$F(3,27) = 0.18$, $p = .912$, BF = 0.01], suggesting that there was no significant difference between correct composite naming scores and **AEFS**. The BF of 0.01 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their correct naming scores.

LIKENESS TASK

Participants rated the composites from 1 (low likeness) to 7 (high likeness) against the target-photograph for likeness. The data were analysed both by participants and by items (see *Likeness Task* section in Chapter 2 for description). It was hypothesised, in agreement with existing research (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984) that providing participants with more opportunity to make RBE (i.e., A>B>C>D) to the facial composite would increase overall composite quality, and, further, the greater the likeness scores would be.

By participants analysis

Likeness scores were analysed by participants. The likeness scores of composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 9.

Table 9: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Participants.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.3	3.4	2.5	2.9
(0.9)	(1.0)	(0.8)	(0.9)

Note: The likeness scale ranged from 1 (low likeness) to 7 (high likeness). For the by participant analysis all group differences were significant, $p < .004$ apart from the test between **AEFS A** and **AEFS B** (p =.67).

There was a significant main effect on mean likeness ratings by **AEFS** [$F(2.63,102.60) = 22.71$, $p < .001$, BF >100]. The BF of > 100 also reveals extreme evidence (e.g., Lee & Wagenmakers, 2014) for the alternative hypothesis. The findings highlight that providing participants with more opportunity to make RBE to the internal facial features of a composite did influence the resulting composites accuracy.

Follow up tests were conducted with a Bonferroni Correction ($\alpha < .05/6$). All paired-samples t-tests (two-tailed) were significantly different ($p < .004$) except between **AEFS A** and **AEFS B** (p =.67). When looking at the results, the hypothesis was partially met. It is evident that composites constructed with less opportunity to make RBE increased the accuracy of the resulting composites, specifically composites constructed in **AEFS B** produced more accurate composites than those constructed in **AEFS C** [with a large ES = 1.1] and **D** [with a medium ES = 0.64]. In terms of the partial hypothesis, it was also important to note that **AEFS D** produced significantly more accurate composites than **AEFS C**, where the composite is usually saved and distributed from in practice [with a large ES = 0.91]**,** supporting the hypothesis that giving witnesses multiple opportunities to make RBE to the internal facial features of a composite did increase its accuracy.

By items analysis

Likeness scores were analysed by items. The likeness scores of composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 10.

Table 10: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Items.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.3	3.4	2.5	2.9
(0.8)	(0.6)	(0.5)	(0.8)

Note: The likeness scale ranged from 1 (low likeness) to 7 (high likeness). Three group differences were significant **AEFS A** and **AEFS C** (*p* = .007), **AEFS B and AEFS C** (*p* = .001) and **AEFS C** and **AEFS D** (*p* = .007).

There was a significant main effect on mean likeness ratings by **AEFS** [$F(3,27) = 6.92$, $p = .001$, BF > 100]. The BF of > 100 also reveals extreme evidence (e.g., Lee & Wagenmakers, 2014) for the alternative hypothesis. The findings highlight that providing participants with more opportunity to make RBE to internal facial features of a composite did influence the resulting composites accuracy.

Follow up tests were conducted with a Bonferroni Correction ($\alpha < .05/6$). The paired-samples t-tests (two-tailed) between **AEFS A** and **C** ($p = .007$), **AEFS B** and **C** ($p = .001$) and **AEFS C** and **D** ($p = .007$) were significant. There was no reliable difference between the remaining paired-samples t-tests ($p > .039$ with an adjusted α level). In more detail, it was evident that composites constructed in **AEFS C**, the point where composites are usually saved in practice, produced significantly less accurate composites than those constructed from **AEFS A** [with a small ES = 0.47], **AEFS B** [with a large ES = 1.5] and **AEFS D** [with a medium ES = 0.76]. Although **AEFS D** did not produce the most accurate composites overall, the composites constructed in this Stage were more accurate than those of **AEFS C**, and so the hypothesis was partially met. These findings were also seen in Experiment 1.

DISCUSSION

It is evident, throughout existing research (e.g., Brown et al., 2017; Chan & LaPaglia, 2011; Fisher & Geiselman, 1992; Fisher et al., 2017; Geiselman et al., 1984; Murre & Dros, 2015; Roediger et al., 2009; Shaw et al., 1995; Wells et al., 2006; Wheeler & Roediger, 1992), that providing witnesses with multiple retrieval attempts at recalling information can improve the accuracy of the information that they provide. However, in Experiment 1 this effect was not observed. The witness-led approach which was adopted in this experiment may have overwhelmed "witnesses" cognitive capacity, preventing them from retrieving reliable information and constructing accurate composites. In regard to the current experiment, it was proposed that providing "witnesses" with multiple opportunities to provide RBE to the internal facial features of a composite, whilst using a practitioner-led approach with a CI, would lead to an overall improvement in the accuracy of the resulting composites. While the percentage of correct naming scores increased from Experiment 1, the results from the current experiment did not support the proposed experimental hypothesis.

The initial hypothesis proposed that providing participants with more opportunity to make RBE to the internal facial features of a composite, would increase the overall percentage of correct composite naming scores was not supported. Findings revealed that there was no reliable difference between the accuracy of composites constructed throughout each **AEFS.** It is important to note though, that correct naming scores did increase from Experiment 1, suggesting that changing the methodology from a witness-led approach to a practitioner-led approach, with the CI, did enhance participants memory, and subsequently, the overall accuracy of composites. These findings align with existing literature, indicating that techniques within the CI (i.e., building rapport, and multiple retrieval attempts) can facilitate witness memory (e.g., Fisher & Geiselman, 1992).

The second hypothesis proposed that providing participants with more opportunity to make RBE to internal facial features, would increase the overall average likeness ratings of the composites. Findings regarding this hypothesis did show a difference across **AEFS**. Contrary to the hypothesis, composites constructed with fewer opportunities to make RBE (**AEFS A** and **AEFS B**) received higher likeness rating scores compared to those constructed with more opportunities (i.e., **AEFS C** and **AEFS D**). A similar trend was observed in the by items analysis, where **AEFS A**, **AEFS B** and **AEFS D** consistently produced composites rated as closer in likeness to the target photograph than those in **AEFS C**, where the composite is usually saved in practice.

The results of the likeness ratings in Experiment 2 support the findings from Experiment 1. In Experiment 1 it was observed that, although **AEFS D** did not consistently produce the most accurate composites, it consistently outperformed **AEFS C**. This trend is consistent with the findings in Experiment 2, suggesting that providing participants with more opportunity to make RBE to the internal facial features of a composite can enhance composite accuracy. This discovery holds significance, especially in practice, as **AEFS C** is where the composite is saved and distributed from when composites are constructed by the police. Over the course of two experiments, it consistently produced the least accurate composites, in terms of the likeness task. These findings are alarming as this suggests that the standard of composites, typically used for identification, may not be of optimum accuracy. The findings across the two experiments have important implications, particularly in light of concerns about composites contributing to wrongful convictions (Innocence Project, 2023). These new findings emphasise the potential benefits of providing participants with more opportunities to make RBE to the internal facial features of a composite, as demonstrated by **AEFS D** when compared to the standard **AEFS C** procedure.

The current experiment's findings do not align with existing research on repeated retrieval methods (e.g., Brown et al., 2017; Chan & LaPaglia, 2011; Fisher & Geiselman, 1992; Fisher et al., 2017; Geiselman et al., 1984; Roediger et al., 2009; Wheeler &. Roediger, 1992). Instead, the findings support alternative research indicating that multiple retrieval attempts do not significantly increase memory accuracy if they are not implemented correctly (e.g., Anderson, 1983; Bjork & Bjork, 1992; Chan & LaPaglia, 2011; Henkel, 2004; Köhnken et al., 1999; Memon & Gawrylowicz, 2006; Memon, Zaragoza et al., 2010; Murre & Dros, 2015; Roediger et al., 1997; Shaw et al., 1995). It is worth considering that these findings might be influenced by the nature of each **AEFS**. **AEFS C** operates on a recognition-based principle, where participants recognise enhancements that closely resemble the offenders face. However, it is unique in that it requires participants to recall information and manipulate facial features in terms of sizing (i.e., thicker/thinner, shorter/wider) and positioning (up, down, left right and rotating) to achieve the optimum likeness towards the target face, a task known to be challenging for individuals (Frowd, Nelson et al., 2012). In contrast, **AEFS D** utilises the Holistic Scales, where participants make holistic judgements about the selected facial features and identify which feature needs enhancement to best resemble the offenders face. Here, participants simply need to choose the feature and recognise its position on the Holistic Scale representing the best likeness to the offender.

Another potential explanation for the **AEFS C** results relates to the methodology of the experiment being conducted online. The procedure heavily relies on communication between participants and the researcher, especially during **AEFS C** when participants need to convey which features to select, and how to adjust their sizing and positioning. It is plausible that communicating these changes over videocall may have been less effective, compared to conducting the experiment face-to-face, where participants could physically point to the screen for feature adjustments.

Although the experimental hypothesis was not met, regarding the increase in the percentage of correct composite naming when comparing across experiments, the change in methodology may be accountable for this. The current experiment adopted a practitioner-led approach, mainly so that the CI could be implemented, and that techniques such as rapport could be applied alongside multiple retrieval attempts (e.g., Chan & LaPaglia, 2011), as multiple retrieval attempts alone did not increase composite accuracy. It is evident that heightened levels of anxiety (e.g., Carter et al., 1996; Eysenck & Calvo, 1992; Fisher & Geiselman, 1992; Kieckhaefer et al., 2014; Vallano & Schreiber Compo, 2011) and cognitive load (e.g., Carter et al., 1996; Eysenck & Calvo, 1992; Kieckhaefer et al., 2014) can overwhelm witnesses processing channels. The modification in methodology, with the researcher leading the session and navigating the software, might have reduced cognitive overwhelm and increased available memory processing channels, permitting participants to focus solely on constructing the composite. Additionally, the use of rapport building techniques before the remainder of the CI and composite construction procedure, could have helped participants feel more at ease aiding in memory access and retrieval. This pattern was consistent with prior research by Frowd, McQuiston-Surrett et al. (2005) who found that the implementation of a CI, produced more accurate results than no interview at all (see *Holistic Interview* section in Chapter 1 for more detail).

This is further supported by the examination of composite accuracy in **AEFS B**. In Experiment 1, **AEBS B** produced the least accurate composites, potentially due to participants struggling, or disengaging with the Stage. The Stage compromises of 15 Holistic Scales, each representing different holistic changes to the face. However, in Experiment 2, **AEFS B** produced the most accurate composites. These findings may be attributed to the practitioner's involvement, specifically as the researchers guided the participants through the procedure and provided verbal instructions, allowing participants to focus on using the scales effectively.

Limitations and future research

A key limitation is that rapport was not directly measured. Rapport is an important component of the CI, but it presents challenges in its definition and measurement (e.g., Richardson & Nash, 2022). However, even without direct measurement, it is plausible that rapport was established, and the beneficial effects (i.e., reduction in anxiety and cognitive load) could have influenced the findings from the experiment, but future research should look to measure this, possibly by asking participants to report what level of rapport they believe had been established (e.g., Richardson & Nash, 2022) after the session.

Another limitation of the experiment is that, despite the change in methodology (witness-led to practitioner-led), certain issues persisted, which could be more controlled in laboratory environments. The practitioner-led version of EvoFIT relied on both the researcher and the participants internet connections. Occasionally, participants experienced connectivity problems, leading to disruptions in the procedure. Furthermore, although it was easier to monitor participant engagement and distractions compared to the witness-led approach, it still presented challenges in maintaining control compared to a controlled laboratory environment (Elliot et al., 2022). These issues could potentially be reduced by conducting the experiment in a face-to-face environment, rather than online, although this depends on the duration of COVID-19 restrictions.

When comparing Experiment 1 and Experiment 2, it is clear that the implementation of the practitioner and the CI did help to increase the overall correct naming of composites. In accordance with existing theory (e.g., Frowd, McQuiston-Surrett et al., 2005) it is known that the CI is known to produce more accurate composites than when no interviewing techniques are implemented. In accordance with this research (e.g., Frowd, Bruce, Smith et al., 2008; Frowd, Nelson et al., 2012) the technique of the H-CI (*see Chapter 1 for more information)* has been shown to increase composite accuracy more so than the CI alone. With this in mind the goal for future research is to incorporate additional memory retrieval techniques known to facilitate memory retrieval, and further, the accuracy of composites. This is important for two reasons. First, it provides insights into how different Stages of EvoFIT function under different conditions (i.e. self-administered procedures, the CI and H-CI). If for instance, it becomes apparent that the Stage in where composites are typically saved and distributed (i.e., **AEFS C**) consistently produces less accurate composites in combination with other investigative interviewing techniques, this issue can be addressed to prevent complications during subsequent composite identification (e.g., Innocence Project, 2023). Second, the implementation of alternative memory facilitating techniques aims to enhance the correct naming of composites, further contributing to the understanding of the different Stages of EvoFIT. Indeed, the finding that Holistic Scales (**AEFS B** and **AEFS D**) have a positive impact on composite accuracy, suggest that it could be valuable to incorporate the H-CI, a memory retrieval technique that helps witnesses to recall information in the same way that it is encoded, holistically (e.g., Frowd, Bruce, Smith et al., 2008), in future studies.

Experiment 2 was conducted to identify whether providing participants with more opportunity to provide RBE to the internal facial features of a composite, would increase the overall accuracy of the resulting composite. Similar to the findings in Experiment 1, the results from Experiment 2 did not support the hypothesis or align with existing literature.

Results suggest that composite accuracy does fluctuate throughout the composite construction procedure. With this in mind, the primary objective of Experiment 3 is to further enhance correct naming scores, aiming to bring them closer to the levels observed in practice (e.g., 60% in Frowd et al., 2019) and previous research studies (e.g., 74% in Frowd et al 2013), whilst still addressing the primary experimental hypothesis. Future research will aim to change the procedure from using a CI to the H-CI. It is evident that the H-CI can be more advantageous than the CI at facilitating memory retrieval and facial recognition (Frowd, Bruce, Smith et al., 2008; Frowd, Nelson et al., 2012), so by changing the interviewing technique, it will hopefully help to facilitate memory retrieval and increase the accuracy of the composites more so than in the current experiment.

CHAPTER 4: EXPERIMENT 3

Asking for more information: Can recognition-based enhancements (RBE) made to internal features produce more accurate composites? *The introduction of a holistic-cognitive interview (H-CI).*

ABSTRACT

The H-CI was developed as an enhancement to the CI, incorporating holistic principles into pre-existing cognitive theories. The H-CI technique leverages insights from how the human brain naturally encodes and recognises faces, making it an important technique for improving memory. The primary objective of Experiment 3 was to investigate whether providing witnesses with more opportunity to make RBE to the internal facial features of a composite, alongside the implementation of the H-CI (e.g., Frowd, Bruce, Smith et al., 2008), would increase the resulting composites accuracy. In Stage 1, 30 participants were individually shown an unfamiliar target face (60 seconds). Twenty-four hours later, participants engaged in a H-CI before constructing a composite using EvoFIT, under one of four experimental conditions. Accumulated EvoFIT Stage **(AEFS) A**: involved selecting a "best face" from face arrays: **AEFS B**: involved **AEFS A**, plus use of Holistic Scales: **AEFS C**: involved **AEFS B**, plus use of the Shape Tool: **AEFS D**: involved **AEFS C**, plus use of Holistic Scales for individual internal features. Stage 2 involved new participants naming (*N* = 28) or rating (*N* = 54) the composites. It was predicted that as the opportunity to make RBE increases (i.e., A>B>C>D) that composite quality should also increase in correlation. In addition to this, it was predicted that the implementation of the H-CI would increase composite accuracy more so to that which is seen in current research and practice. Correct composite naming scores were 23.9% with no reliable difference between **AEFS**. Regarding the likeness task, there was one reliable finding for the external features: by participants analysis. It was found the composites constructed in **AEFS A** produced more accurate composites suggesting that providing participants with more opportunity to make RBE did not produce more accurate composites.

INTRODUCTION

The process of constructing a composite can be complex, primarily due to the challenges associated with encoding information at the crime scene (e.g., facial details of the offender) and later retrieving that information during subsequent recall opportunities, or recognising facial details when constructing a composite (Frowd, Bruce & Hancock, 2008; Frowd, Bruce, Smith et al., 2008). It is important to recognise that if the process of encoding and decoding information is disrupted, it can have detrimental effects on the retrieval of accurate memories (Baddeley, 1979). This, in turn, can compromise the accuracy of composites. In some cases, the use of such inaccurate composites can even lead to the wrongful arrest of innocent individuals, as observed in cases reported by organisations such as the Innocence Project (2023).

Therefore, the primary objective of the current experiment is to improve witnesses encoding and retrieval of information about an unfamiliar face, with the ultimate goal of enhancing the accuracy of resulting composite images to a more reliable level.

Memory retrieval

The accuracy of retrieving facial information has been a subject of past research, revealing that both encoding and decoding processes significantly impact the quality of retrieved information (e.g., Fisher & Geiselman, 1992; Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd, McQuiston-Surrett et al., 2005; Geiselman et al., 1984; see Köhnken et al., 1999 for a meta-analysis). Witnesses play a critical role in providing evidence, relying on their ability to accurately recall memories (Frowd et al., 2004; Frowd, Hancock et al., 2011; Frowd et al., 2013) to retrieve information about an offender's face in order to construct a composite of it (Brace et al., 2006). It is important to acknowledge that various factors can influence both face processing (e.g., inspection period; how long the face was observed for) and the recall of facial features (e.g., how the face was processed and retrieval intervals; the time between observing the face and recalling details).

Recalling specific facial features can be quite challenging due to the requirement of using multiple cognitive resources to ensure effectiveness and accuracy (Frowd, Bruce, Smith et al., 2008). This task is often perceived as demanding, and time-consuming, with the added complication of memories fading relatively quickly. Due to the intricacies involved in retrieving facial memories, errors can occur during subsequent face recognition tasks (see, Meissner & Brigham, 2001 for meta-analysis). This is evident in various scenarios, such as recognising an offender’s face in a line-up, or identifying features when constructing a composite.

While many investigative interviews (e.g., the CI), and composite systems (e.g., E-FIT, FACES and PRO-fit) traditionally focus upon the recall of individual facial features to obtain descriptions, extensive research indicates that faces are in fact processed holistically (Bruce & Young, 1998; Richler et al., 2009). This holistic approach contrasts with the conventional interviewing techniques, and composite systems, that rely heavily on describing and identifying individual facial characteristics. The difference in processing may contribute to the difficulties witnesses face when trying to recognise a face at a later point, or determine whether a composite accurately represents the original face (Frowd, Nelson et al., 2012; Wells & Hryciw, 1984). When a holistic approach has been applied to memory retrieval procedures (i.e., investigative interviewing) it has been found to substantially enhance memory recall and, consequently, the correct recognition and identification of familiar faces (e.g., Bruce & Young, 1998; Fodarella et al., 2015; Frowd, Bruce & Hancock, 2008; Frowd, Bruce, Smith et al., 2008; Frowd et al., 2004; Frowd et al., 2013; Wells & Hryciw, 1984).

Verbal overshadowing effect (VOE)

In the context of investigative interviewing and composite construction, it becomes evident that two fundamentally different cognitive processes (Wells & Hryciw, 1984) are utilised: recall and facial recognition. Recall involves the task of retrieving and describing individual facial features from memory, but is a time-consuming task, and memories tend to decay rapidly over time. On the other hand, facial recognition is characterised by its speed, accuracy, and stability over time (Frowd, Bruce, Smith et al., 2008) which as a result enhances witness memory (Frowd, Bruce, Ness et al., 2007; Frowd, McQuiston-Surrett et al., 2005).

The encoding specificity principle (Tulving & Thomson, 1973) suggests that accurate memory retrieval relies on using the same method for both encoding and decoding of information. To illustrate, if a face is processed holistically, as research suggests, it is proposed to use recognition as the retrieval method, as switching from holistic processing to retrieving individual facial features may increase the chance of errors.

Related to this is the verbal overshadowing effect (VOE; see Schooler & Engstler-Schooler, 1990 and *Verbal Overshadowing Effect* section in Chapter 1 for detail). The VOE explains that errors in eyewitness memory could be due to a shift between hemispheres in the brain, during encoding and decoding information. The shift may occur when processing a face holistically (based in the right hemisphere) and then trying to retrieve individual facial features (based in the left hemisphere), potentially leading to errors (Frowd, Bruce, Smith et al., 2008; Wells & Hryciw, 1984). Studies have observed that when witnesses provided a facial description of a face that they have seen, they were more likely to misidentify the face when it came to recognition at a later point (Schooler & Engstler-Schooler, 1990; see Meissner & Brigham, 2001 for a meta-analysis). This discovery aligns with the findings of Schooler and Engstler-Schooler (1990), who conducted studies revealing negative consequences for recognition when witnesses were asked to recall and provide a verbal description of individual facial features.

There are contrasting views on the relationship between encoding and decoding methods in facial recognition processes. Nairne (2002) proposes that the link between these methods (i.e., holistic processing and recognition vs feature-based processing and recall of facial details) is primarily correlational, instead of causal, and that other factors are in fact responsible for the effect of VOE. Frowd and Fields (2011) conducted research indicating that VOE only manifests when there is no delay between describing the target face and constructing a composite image. In situations where there is a delay between retrieving facial details and constructing a composite, or when no verbal description was provided at all, the VOE is not observed. It is worth noting that introducing a short delay (e.g., 24 minutes; Finger & Pezdek, 1999) can help to "release" the VOE. However, this may be difficult to achieve in practice with limited resources (e.g., Fodarella et al., 2015; Frowd, Pitchford et al., 2012; Frowd, Skelton et al., 2012). This suggests that the brief pause before recognition might prevent forgetting and potentially reduce the impact of the VOE (see *Retention Interval* section in Chapter 1 for more detail particularly the concept of the forgetting curve; Murre & Dros, 2015). This implies that the VOE may be mitigated through the strategic introduction of a short delay in the recall process.

Holistic processing

Investigative interviewing processes, and composite systems, often emphasise the importance of recalling individual facial details. However, research has revealed a fundamental shift in the understanding of how faces are processed. Instead of processing faces by focusing on specific features (e.g., eyes, nose, and mouth) it has become clear that faces are primarily processed holistically, that is as a whole (Bruce & Young, 1998; Richler et al., 2009).

This may explain why feature-based computer software and feature-focused interviews (e.g., the CI) generally tend to produce less accurate results compared to recognition-based (or holistic) composite software's and interviews (e.g., the HI and H-CI)

(Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd et al., 2015; Frowd, McQuiston-Surrett et al., 2005; Frowd, Nelson et al., 2012). The underlying issue lies in the mismatch between the encoding and decoding methods employed in the two approaches. When a witness tries to construct a composite using feature-based software, based on the recall of individual facial features, if they originally perceived the face holistically, this task becomes challenging (Schooler & Engstler-Schooler, 1990).

When investigating the concept of holistic processing in the context of face recognition, it becomes evident that the recognition of a face is typically enhanced when it is encoded holistically, with a focus on character attributes, rather than solely relying on physical attributes (Shapiro & Penrod, 1986; Wells & Hryciw, 1984; Winograd, 1976, 1981). Consequently, it has been recognised that incorporating a holistic approach in both the interviewing procedure and composite construction process can enhance memory retrieval in witnesses (Frowd, Bruce, Smith et al., 2008). This recognition of the advantages of holistic processing in facial recognition played a vital role in the development of the HI.

In light of this connection an alternate approach to the traditional CI (which primarily relies on recalling individual features) known as the HI was developed (Frowd, Bruce, Ness et al., 2007; Frowd, McQuiston-Surrett et al., 2005). The HI draws its theoretical foundations from the work of Berman and Cutler (1998), who discovered that recognising a face can be enhanced when individuals consider holistic aspects such as the personality traits associated with the face. In practical terms, the HI involves witnesses reflecting on the face, and then making judgements about the personality attributes (Frowd, Nelson et al., 2012). To parallel the CI's focus on seven physical attributes, the HI incorporates a similar structure by asking participants to rate seven personality attributes (e.g., aggression, hostility, and intelligence) on a scale of low, medium, or high.

One important study that led to the development of the HI (see Frowd, McQuiston-Surrett et al., 2005) involved participants viewing an unfamiliar face. After 48 hours they were then subject to either no interview, a CI, or a HI, before constructing a composite. The findings from this study revealed an advantage in the quality of composites constructed through the HI, exceeding those constructed with either no interview, or a CI. Subsequently, in a similar study, conducted by Frowd, Bruce, Ness et al. (2007) participants observed a video featuring an unfamiliar identity. Participants were then interviewed using either the CI or the HI. Again, the findings from this study found an approaching benefit for composites constructed using a HI, supporting existing research (e.g., Berman & Cutler 1998; Shapiro & Penrod, 1986; Wells &

Hryciw, 1984) which highlight the memory enhancing effects of making holistic judgements about faces.

However, it became evident that the HI faced difficulties when applied in conjunction with a feature-based system (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, McQuiston-Surrett et al., 2005) given that the systems nature involves selecting individual facial features to create an initial face (Frowd, Bruce, Smith et al., 2008). Despite this limitation, the HI still continued to produce more accurate composites than the CI. Recognising the potential of the HI and its ability to improve composite quality (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, McQuiston-Surrett et al., 2005), Frowd, Bruce, Smith et al. (2008) used these findings to develop a more versatile interviewing technique, suitable for a range of systems. They developed a "hybrid" procedure known as the H-CI, which merged components from both the CI and HI (Frowd, Bruce, Smith et al., 2008).

The H-CI primarily involves witnesses engaging in a CI (building rapport and recalling facial details, through multiple opportunities). The CI component is important as it permits witnesses to reflect on the face and recall important information. The HI component invites witnesses to think about the personality of the face and rate personality traits (e.g., attractiveness, aggressiveness, hostility, and intelligence) (Frowd, Bruce, Smith et al., 2008). The HI perspective is believed to enhance the recognition of familiar features as it works in the same way as face processing.

Frowd, Bruce, Smith et al. (2008) conducted a study to assess the new "hybrid" interviewing procedure. This study involved participants observing a video containing an unfamiliar identity, and then being interviewed three to four hours later, using either a CI or the new H-CI method. The key finding from this study was that participants who engaged in a H-CI produced composites that were correctly identified four times more frequently than those interviewed using the standard CI. Moreover, there was also a notable reduction, of over 50%, in the incorrect naming percentages of composites when the H-CI was employed.

This suggests that implementing a H-CI into the investigative interviewing process helps to produce more accurate composites, compared to using the CI isolated. Supporting this, and disputing previous research (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, McQuiston-Surrett et al., 2005), Frowd, Nelson et al. (2012) discovered that the recall component (CI) of the H-CI produced more accurate composites than the HI alone. However, the most accurate composites were constructed when both the CI and HI were combined as the H-CI. Overall, the research predicted that if the H-CI was to be used in practice it would significantly enhance identification of offenders (Frowd, Nelson et al., 2012) and potentially reduce the number of wrongful convictions (i.e., Innocence Project, 2023). This advancement in composite accuracy could have a direct impact on improving the outcomes of the **AEFS** by providing more reliable data for analysis.

Current research

In summary, it is clear that a holistic approach to encoding facial information, in combination with the enhanced retrieval of facial details can increasingly improve witness memory when it comes to investigative interviews and composite construction. It is hypothesised that implementing the H-CI (e.g., Frowd, Bruce, Smith et al., 2008) into the current experiment will not only enhance witnesses' memory, but also improve the accuracy of the resulting composites. Furthermore, it is expected that the implementation of the H-CI will positively influence the new Holistic Scale (**AEFS D**), as existing research has shown that holistic judgements play a crucial role in the facilitation of face recognition (Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd, McQuiston-Surrett et al., 2005; Frowd, Nelson et al., 2012). Overall, it is predicted that the more opportunity that participants are given to make RBE to the composite (i.e., A>B>C>D), the more accurate the composite will be, as compared to those composites constructed with limited opportunity to do so.

METHOD

As with the previous experiments, Experiment 3 was divided into 2 Stages. Stage 1 involved the procedure of constructing a composite of an unfamiliar face. Stage 2 involved a naming and likeness task to evaluate the accuracy of the composites constructed in Stage 1.

Ethics

Ethical approval was received from the ethics board within the School of Psychology at UCLan.

Design

The design for the experiment is the same as that in both Experiment 1 and 2.

STAGE 1: COMPOSITE CONSTRUCTION

Participants

Thirty participants, 17 females and 13 males (*M* = 29.6, *SD* = 12.1) who were unfamiliar with the television soap EastEnders, were recruited via opportunity sampling. Participants were recruited in the same way as the previous experiments. Participants recruited through SONA received 7 SONA points for their participation, whilst those recruited through recruitment websites received a £5 Amazon voucher.

Materials

The target-photographs were selected in the same way as the previous two experiments, via an extensive search of the internet by Professor Frowd. The target-photograph identities were actors and actresses from the television soap EastEnders. The identities were Ian Beale, Jane Beale, Jack Branning, Lauren Branning, Max Branning, Shirley Carter, Martin Fowler, Billy Mitchell, Jean Slater, and Stacey Slater. The target-photographs were 8cm x 10cm and in colour.

Procedure

Day one of the experiment followed the same procedure as the previous experiments whereby the participant individually looked at a target-photograph of an unfamiliar identity for 60 seconds with the aim of remembering facial details.

Twenty-four hours later, the participant was invited back to construct a composite of the target face. Before constructing the composite, a H-CI (e.g., Frowd, Bruce, Smith et al., 2008) was administered. The CI component (e.g., Fisher & Geiselman, 1992), as with Experiment 2, involved building an informal rapport with the participant before asking them to think about the face (60 seconds) that they saw 24-hours prior. After thinking about the face, the participant freely recalled, without interruption, everything that they could remember about the face. The CI component of the session took 15 minutes to complete and was conducted in the same way as Experiment 2. Once the CI component had been administered, the participant then completed the holistic component (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd, McQuiston-Surrett et al., 2005) of the interview. This involved the participant thinking about the personality of the whole face for 60 seconds. Following this, the participant was then requested to make several personality judgements (e.g., attractiveness, aggressiveness, and honesty) about the face, on a scale of; low, medium, or high. In practise, these judgements are chosen based around the circumstances of the crime (e.g., aggression and hostility for a violent crime). After this, the participant was asked to think about (60 seconds) the eye region (as the initial Stage of EvoFIT, **AEFS A**, involves choosing faces based on the eye region) before repeating the task of rating several personality judgements for the region. This task took around 10 minutes to complete.

Once the full H-CI had been administered the researcher then opened the EvoFIT software and shared their screen with the participant. Together, the participant and the researcher worked through the software to construct a composite. The procedure of composite construction was the same as the practitioner-led version in Experiment 2 (see *Procedure* section in Chapter 3 for in-depth procedure). Participants were randomly allocated to one of the four conditions (**AEFS A**, **B**, **C**, or **D**).

STAGE 2: COMPOSITE EVALUATION

NAMING TASK

The naming task involved participants looking at an array of composites and attempting to identify them.

Participants

Twenty-eight participants, 18 females and 10 males (*M* = 36.2, *SD* = 13.9) who were familiar with the television soap EastEnders, were recruited in the same way as the previous experiments, via opportunity sampling. Participants who were recruited through SONA received 1.5 SONA points for their participation, where participants recruited through recruitment websites received a £2 Amazon voucher.

Materials and Procedure

The materials and procedure for the naming task were the same as in previous experiments. As part of the naming task, an *a priori* rule was specified; participants needed to recognise 80% of the ten target-photographs to have their data included in the analysis. All participants met this specification and therefore no data were excluded from the analysis. There was also no missing data.

LIKENESS TASK

The likeness task involved participants rating the accuracy of composites against the target face on a Likert scale of 1 (low likeness) to 7 (high likeness).

Participants

Fifty-four participants, 21 females and 33 males (*M* = 28.4, *SD* = 12.2), who were unfamiliar with the television soap EastEnders were recruited in the same way as the previous experiments, via opportunity sampling. Participants recruited through SONA received 1.5 SONA points for their participation, where those recruited through recruitment websites received a £2 Amazon voucher.

Materials and Procedure

The materials and procedure for the likeness task were the same as in the previous experiments, with one exception. Experiment 3 had participants rate the likeness of the constructed composites towards to target face, not only the whole face (see Figure 11 for an example), but also for either internal facial features (e.g., eyes, nose, and mouth; see Figure 12 for an example) or external features (e.g., hair, ears, and face shape; see Figure 13 for an example). This new task was completed as it is evident that there are differences between internal and external features (Frowd, Bruce, McIntyre et al., 2007) in that familiar faces are more recognised by their internal facial features (e.g., Ellis et al., 1979). Participants were randomly allocated to either one of three conditions.

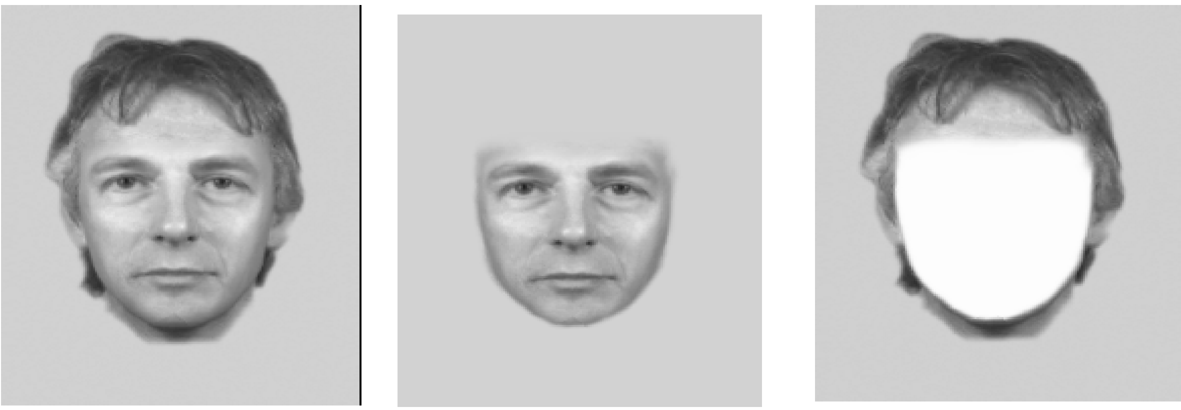


Figure 11. An example EvoFIT composite for whole face features (left). Figure 12. An example EvoFIT composite for internal facial features (middle). Figure 13. An example EvoFIT composite for external features (right).

RESULTS

ANALYSIS

The traditional ANOVA is adopted throughout the thesis to analyse the data. However, the traditional ANOVA does not provide the evidence needed for both the null and alternative hypothesis; therefore, Bayesian Factors ANOVA will be adopted throughout the thesis in order to provide that information.

MANIPULATION CHECK

Target-photographs were named very well (*M* = 99.3, *SD* = 3.8) with minor differences between **AEFS** meaning that all participants had the same potential to correctly identify all composites in each **AEFS**. The naming scores are summarised in Table 11.

Table 11: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Target-Photograph Naming.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
97.1	100.0	100.0	100.0
(7.6)	(0.0)	(0.0)	(0.0)

As a manipulation check, correct target-photograph naming scores by **AEFS** were analysed using an independent samples ANOVA. There was a non-significant main effect on correct target-photograph naming scores by **AEFS** [$F(3,24) = 1.00$, $p = .410$, BF = 0.04]. The BF of 0.04 also reveals strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. This suggests that there was no significant difference between correct target-photograph naming scores by **AEFS**.

NAMING TASK

Responses were scored for accuracy. A score of 1 was given for correct responses and 0 if no name, or an incorrect name was provided. Participants responses were checked for missing data, of which non was found. Overall, correct composite naming scores were 23.9% (*SD* = 13.4). The naming data were analysed both by participants and by items (see *Naming Task* section in Chapter 2 for description). It was hypothesised, in agreement with existing research (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984) that providing participants with more opportunity to make RBE (i.e., A>B>C>D) to the facial composite would increase overall composite quality, and, further, the greater the percentage of correct composite naming scores would be.

By participants analysis

Correct composite naming percentages were analysed by participants. The percentage of correctly named composites, by **AEFS**, was analysed using an independent samples ANOVA. An independent samples Bayesian Factors ANOVA was also conducted to find the BF value. The naming scores are summarised in Table 12.

Table 12: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Composite Naming: By Participants.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
25.7	21.4	28.6	20.0
(5.3)	(10.7)	(17.7)	(17.3)

There was a non-significant main effect on correct composite naming scores by **AEFS** [$F(3,24) = 0.57$, $p = .639$, BF = 0.02], suggesting that there was no significant difference between correct composite naming scores and **AEFS**. The BF of 0.02 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their correct naming scores.

By items analysis

Correct composite naming percentages were analysed by items. The percentage of correctly named composites, by **AEFS**, was analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The naming scores are summarised in Table 13.

Table 13: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Composite Naming: By Items.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
18.0	15.0	19.1	14.0
(22.5)	(19.0)	(20.7)	(16.5)

There was a non-significant main effect on correct composite naming scores by **AEFS** [$F(3,27) = 0.20$, $p = .896$, BF = 0.01], suggesting that there was no significant difference between correct composite naming scores and **AEFS**. The BF of 0.01 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their correct naming scores.

LIKENESS TASK

Participants rated the composites, on a Likert scale, of 1 (low likeness) to 7 (high likeness) for likeness to the target-photograph.

Likeness scores were analysed for the constructed composites; whole face (both internal and external features), internal features (e.g., eyes, nose, and mouth) and external features (e.g., hair, ears, and face shape). Analysis was also conducted both by participants and by items (see *Likeness Task* in Chapter 2 for description). It was hypothesised, in agreement with existing research (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984) that providing participants with more opportunity to make RBE (i.e., A>B>C>D) to the facial composite would increase overall composite quality, and, further, the greater the likeness scores would be.

<u>WHOLE FACE ANALYSIS</u>

By participants analysis

Likeness scores were analysed, by participants, for the whole face of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 14.

Table 14: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Participants for Whole Face.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.7	3.6	3.5	3.6
(1.3)	(1.2)	(1.1)	(1.2)

There was a non-significant main effect on likeness ratings by **AEFS** [$F(3,51) = 0.66$, $p = .584$, BF = 0.03], suggesting that there was no significant difference between likeness scores and **AEFS**. The BF of 0.03 also reveals strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their likeness scores.

By items analysis

Likeness scores were analysed, by items, for the whole face of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 15.

Table 15: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Items for Whole Face.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.7	3.6	3.5	3.6
(0.7)	(0.7)	(0.7)	(0.8)

There was a non-significant main effect on likeness ratings by **AEFS** [$F(3,27) = 0.16$, $p = .919$, BF = 0.01], suggesting that there was no significant difference between likeness scores and **AEFS**. The BF of 0.01 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this suggests that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composite or their likeness scores.

INTERNAL FEATURES ANALYSIS

By participants analysis

Likeness scores were analysed, by participants, for the internal facial features of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 16.

Table 16: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Participants for Internal Features.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.2 (0.9)	3.2 (0.9)	3.5 (0.9)	3.5 (1.0)

There was a non-significant main effect on likeness scores by **AEFS** [$F(3,51) = 2.6$, p =.063, BF = 0.47], suggesting that there was no significant difference between likeness scores and **AEFS**. The BF of 0.47 also reveals anecdotal evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their likeness scores.

By items analysis

Likeness scores were analysed, by items, for the internal facial features of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 17.

Table 17: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Items for Internal Features.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.2	3.2	3.5	3.5
(0.6)	(0.9)	(0.7)	(0.7)

There was a non-significant main effect on likeness ratings by **AEFS** [$F(2.1,19.0) = 1.10$, $p = .356$, BF = 0.09], suggesting that there was no significant difference between likeness scores and **AEFS**. The BF of 0.09 also reveals strong evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their likeness scores.

<u>EXTERNAL FEATURES ANALYSIS</u>

By participants analysis

Likeness scores were analysed, by participants, for the external facial features of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 18.

Table 18: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Participants for External Features.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
4.6	4.3	4.1	4.1
(1.1)	(0.9)	(0.8)	(0.8)

Note: The likeness scale ranged from 1 (low likeness) to 7 (high likeness). For the by participants analysis two group differences were significant **AEFS A** and **AEFS C** ($p = .004$) **AEFS A** and **AEFS D** ($p = .014$).

There was a significant main effect on likeness scores by **AEFS** [$F(3,51) = 5.37$, $p = .003$, BF = 19.13]. The BF of 19.13 also reveals strong evidence (e.g., Lee & Wagenmakers, 2014) for the alternative hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to internal facial features of a composite did influence the accuracy of composites and likeness scores.

Follow up tests were conducted with a Bonferroni Correction ($\alpha < .05/6$). Two of the six paired-samples t-tests (two-tailed) revealed a significant difference between composite likeness scores and **AEFS** (**AEFS A** and **AEFS C**, $p = .004$; **AEFS A** and **AEFS D**, $p = .014$). All remaining t-tests were non-significant ($p > .071$). In more detail, **AEFS A** produced composites that were rated as significantly more accurate than **AEFS C** [with a medium ES = .78] and **AEFS D** [with a medium ES = 0.64]. This suggests that the external facial features of the composites, that were constructed with a lower opportunity to make RBE, produced more accurate composites than those created with greater opportunity.

By items analysis

Likeness scores were analysed, by items, for the external facial features of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 19.

Table 19: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Items for External Features.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
4.6	4.3	4.1	4.1
(0.5)	(0.6)	(1.2)	(1.1)

There was a non-significant main effect on likeness ratings by **AEFS** [$F(3,27) = 1.10$, $p = .317$, BF = 0.14], suggesting that there was no significant difference between likeness scores and **AEFS**. The BF of 0.14 also reveals moderate evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their likeness scores.

DISCUSSION

In many investigative interviews (e.g., the CI; Fisher & Geiselman, 1992; Geiselman et al., 1984) and composite systems (e.g., E-FIT, FACES and PRO-fit) the conventional approach to face recall has been to focus on recalling individual facial features. However, a substantial body of literature on face processing suggests that the human brain processes a face holistically, rather than by isolated features (Bruce & Young, 1998; Richler et al., 2009). This means that, to achieve the most optimum results when retrieving information about a face, the retrieval method should align with holistic processing (Schooler & Engstler-Schooler, 1990). Therefore, it is suggested that holistic techniques, such as forming personality judgements based on facial characteristics, (e.g., Shapiro & Penrod, 1986; Wells & Hryciw, 1984; Winograd, 1976, 1981) should be employed during investigative interviews to obtain more accurate and reliable information.

The aim of the current experiment was to identify whether providing participants with more opportunity to make RBE to the internal facial features of a composite would increase the overall accuracy of the resulting composites alongside the implementation of a H-CI. The hypothesis was not met.

The initial hypothesis of the experiment proposed that providing participants with more opportunity to make RBE to the internal facial features of a composite would increase the overall percentage of correctly named composites. However, this hypothesis was not supported by the findings. In more detail, the results revealed no reliable difference between the accuracy of composites constructed across each **AEFS**. It is important to note that although the current experiment employed a holistic approach during the interviewing process, unlike previous research (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd, McQuiston-Surrett et al., 2005; Frowd, Nelson et al., 2012), this approach did not result in an overall increase in the percentage of correctly named composites. This outcome was unexpected, as it was hypothesised that the implementation of the H-CI would improve facial recognition, and as a consequence, would have increased composite accuracy. Surprisingly, the percentage of correctly named composites actually decreased in the current experiment compared to Experiment 2, with a reduction of 3.1%.

Hypothesis 2 proposed that providing participants with more opportunity to make RBE to the internal facial features of a composite would increase the overall composite likeness ratings. However, results revealed that there was no reliable difference between the accuracy of composites constructed across each **AEFS**, except for one analysis. Specifically, the analysis focusing on external features, as assessed by participants, revealed that composites constructed with fewer opportunities (**AEFS A**) to make RBE to the internal facial features of a composite produced more accurate composites.

The lack of reliable difference between composite accuracy throughout the **AEFS**, with composites presenting similar levels of accuracy, can potentially be attributed to the implementation of a holistic interviewing approach (the H-CI) within the procedure. In this case, it suggests that if both the encoding and retrieval of facial information employ a holistic approach, memory performance is likely to be optimised, and composites constructed are likely to be more of an accurate representation of the offenders' face (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd et al., 2015; Frowd, McQuiston-Surrett et al., 2005; Frowd, Nelson et al., 2012). In the present experiment, it is evident throughout both naming and likeness analyses (apart from external features analysed by participants) there was no reliable difference in composite accuracy amongst the different **AEFS**. Considering that facial processing predominantly relies on holistic perception, rather than individual features (Frowd, Bruce, Smith et al., 2008), the incorporation of a holistic aspect into the interviewing procedure might have played a pivotal role. It is possible that the H-CI approach assisted participants in constructing composites of comparable accuracy across all **AEFS**, effectively acting as a ceiling effect and minimising differences in accuracy. While the implementation of the H-CI did not necessarily lead to an overall improvement in composite accuracy, it likely facilitated participants memory process to such an extent that differences between **AEFS** became unidentifiable. These findings hold incredible benefits for forensic practitioners. Notably, the experiments so far in this thesis reveal that the implementation of no interview (Experiment 1), and the CI (Experiment 2) resulted in varying levels of composite accuracy across all **AEFS**. In contrast, the H-CI approach consistently produced stable composite accuracy across all **AEFS**. This suggests that integrating a H-CI interviewing approach can lead to more consistent outcomes, in regard to composite accuracy, regardless of whether participants are provided more opportunity to make RBE.

The pattern of results could be explained by referring back to the information presented in Chapter 2, it is important to consider the possibility that the outcomes discovered in this experiment could be attributed to retrieval induced forgetting (MacLeod, 2002; Shaw et al., 1995). This is when information related to a particular subject, in this case information about a face, is not successfully retrieved during the initial recall attempt (i.e., in the case of the current experiment, during the CI) it may become increasingly difficult to recall that information in subsequent attempts. In this case, it may be that when participants were provided with additional opportunities to make RBE to the internal facial features of the composite, it did not necessarily facilitate the recall of new information and instead, led to participants repeatedly recalling the same information about the face. This might be why, despite providing participants with more opportunity to recall information in the interviewing procedure, similar outcomes for accuracy were found throughout the **AEFS**.

In the context of witness memory, and face recognition, it is important to consider the effect of VOE. In Experiment 2, the study involved both holistic processing of the target face, and individual feature recall (through the CI). Surprisingly, this approach resulted in more accurate composite images compared to the current experiment, where holistic processing, holistic judgements and a holistic system were used. This is intriguing, as typically, making holistic personality judgements are known to increase composite accuracy (Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd, McQuiston-Surrett et al., 2005; Frowd, Nelson et al., 2012). One possible explanation for this finding is that, in both Experiments 2 and 3, participants were made aware that they would need to recall facial details at a later point, to construct a composite. This awareness, in support of Laughery et al. (1986) seems to have aided the accurate retrieval of facial details, as participants likely paid close attention to these features during the initial encoding process. However, participants were not informed about the need to consider personality attributes until it came to interview. As a result, they may not have given as much thought to the personality attributes during the initial encoding of the face, focusing instead on individual features.

Limitations and future directions

One limitation of the experiment was discrepancy in sample sizes between the previous experiments and the current one. This discrepancy primarily arose from opportunity sampling. In Experiment 3, recruiting participants for the naming task proved to be more challenging when compared to Experiment 1 and 2. This difficulty may be attributed to the choice of television soap, EastEnders, which potentially has a smaller viewership. It is important to note that a large target audience for the research is university students. However, Eastenders, and other television soaps for that matter, may not align with the viewing interests of this demographic. With the increase of streaming services, offering a variety of content, university students may be less inclined to watch traditional television soaps. Conversely, the previous two experiments, which used Coronation Street as the target identities, were easier to recruit for. Coronation Street is set in the north-west of England, which is on close proximity to the university. This geographical proximity may have increased participants familiarity with Coronation Street. Given the recruitment difficulties with EastEnders, it may be beneficial to reconsider using Coronation Street as target identities in future experiments.

Another aspect to consider is the environment in which the experiment was conducted. Experiment 3 relied on online data collection, which inherently limits the control that researchers have over participants surroundings (Elliot et al., 2022). Each participant is tasked with encoding facial features, under varying conditions (e.g., different size laptop screens, or see the face under different lighting i.e., on the laptop or lighting in the room). In contrast, traditional laboratory experiments typically have a controlled environment in where participants complete the experiment under the same conditions. To overcome this, it would be beneficial to change to face-to-face testing now that COVID-19 restrictions are lifting. However, individuals still may have concerns about face-to-face interactions, and this could potentially impact recruitment.

As face-to-face sessions are not really permissible at the moment, it is important to address other factors that can be implemented to help increase the accuracy of constructed composites. In light of this, a factor that has been highlighted throughout this thesis as enhancing information retrieval, is the retention interval (e.g., Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; see Frowd et al., 2015 for a meta-analysis). Although memory facilitating techniques such as the CI (Fisher & Geiselman, 1992; Geiselman et al., 1984) and the H-CI (Frowd, Bruce, Smith et al., 2008), which have been tested throughout this thesis, are known to increase memory accuracy in witnesses, memory decay can still be detected when these facilitating techniques are applied after 24 hours (Frowd, Carson, Ness, McQuiston-Surrett et al., 2005). By reducing the retention interval between the witness inspecting the target-photograph and constructing the composite, it should, as literature suggests, enhance recall and recognition of a face at a later point and reduce factors such as VOE and forgetting (Frowd & Fields, 2011; Murre & Dros, 2015).

Experiment 3 aimed to identify if providing participants with more opportunity to make RBE to the internal facial features of a composite, whilst employing a H-CI, would enhance the accuracy of the resulting composites. The findings from Experiment 3 did not support the experimental hypothesis or align with previous research on this topic (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd, McQuiston-Surrett et al., 2005, Frowd, Nelson et al., 2012).

Considering these results, Experiment 4 will focus on reducing the retention interval, between the inspection period and composite construction, from 24 hours to a shorter duration, ranging between 3-4 hours. This adaption aims to maintain the focus on the experimental hypothesis, whilst hopefully increasing the percentage of correct composite naming scores.

CHAPTER 5: EXPERIMENT 4

Asking for more information: Can recognition-based enhancements (RBE) made to internal features produce more accurate composites? *The introduction of a short retention interval (3-4 hours).*

ABSTRACT

The retention interval between a witness observing a criminal offence and either retrieving information about the offenders' face, or constructing a composite of it, has been shown to impact the accuracy of information gathered. In practice, these retention intervals can be days, or even weeks, after a crime has taken place. This is problematic given that research suggests the longer the delay, the greater the room for misinformation. However, it is possible that multiple retrieval opportunities, after a short delay, can help reduce these errors (Hintzman & Rogers, 1973; Murre & Dro's, 2015; Roediger et al., 2009). In Stage 1 of the experiment, 30 participants looked at an unfamiliar target face (for 60 seconds). Three to four hours later participants were then invited to construct a composite using EvoFIT. Participants were placed into one of four Stages, representing different levels of opportunity to make RBE to a composite, known as Accumulated EvoFIT Stages: **(AEFS) A**: involved selecting a "best face" from face arrays: **AEFS B**: involved **AEFS A**, plus use of Holistic Scales: **AEFS C**: involved **AEFS B**, plus use of the Shape Tool: **AEFS D**: involved **AEFS C**, plus use of Holistic Scales for individual internal features. In Stage 2, new participants were asked to either name (N = 28) or rate (N = 54) the composites. It was predicted that as the opportunity to make RBE increases (i.e., A>B>C>D) that composite quality should also increase in correlation. In addition to this, it was hypothesised that implementing a shorter retention interval would increase composite accuracy to levels seen in current research and practice. Findings supported the hypothesis that the more opportunity to make RBE to internal facial features of a composite the greater the accuracy of the resulting composite.

INTRODUCTION

In the context of criminal procedures, it is standard practice for witnesses to provide a detailed and accurate description of an offender's facial features (Frowd et al., 2004; Frowd, Pitchford et al., 2012). However, the process of recalling this key evidence can occur over several occasions spanning a considerable period of time (Brown et al., 2020; Tukey & Brewer, 2003). The first opportunity for witnesses to recall and provide evidence typically arises during an investigative interview, or when constructing a composite image of the offenders' face. Importantly, this initial retrieval opportunity may occur several days, or even weeks, after the crime has taken place (Fodarella et al., 2015; Frowd, Pitchford et al., 2012; Frowd, Skelton et al., 2012; Roediger & Marsh, 2009; Tukey & Brewer, 2003). The time passed between witnesses observing the crime, and retrieving information about it, is known as the retention interval, a concept which can systematically impact the accuracy of recalled information (Krouse, 1981; Shepherd et al., 1991). Shorter retention intervals have been found to be advantageous as they increase the proximity between the initial encoding of information (i.e., seeing an offender's face) and the subsequent retrieval or recognition of the offenders' facial features (Brown et al., 2020; Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Frowd, Carson, Ness, Richardson et al., 2005; Frowd et al., 2015). Conversely, longer retention intervals between encoding and retrieval of information can result in information being recalled at a slower rate, requiring superior cognitive effort to recover (Frowd, Carson, Ness, McQuiston-Surrett et al., 2005). However, it is important to note that effective retrieval practices can mitigate the negative impact of longer retention intervals on memory accuracy (e.g., Butler & Roediger, 2007; Murre et al., 2015; Odinot & Wolters, 2006; Roediger & Karpicke, 2006; Roediger et al., 2009).

The current experiment aims to test whether providing participants with more opportunity to make RBE to the internal facial features of a composite would increase the composites overall accuracy, specifically with a short retention interval ranging between 3-4 hours.

The impact of retention intervals

In practice the retention interval, which is the time elapsed between an individual witnessing a crime and engaging in an investigative interview, typically ranges between 24-48 hours. However, reducing this interval has been deemed challenging due to constraints such as limited police resources and availability (e.g., Fodarella et al., 2015; Frowd, Pitchford et al., 2012;

Frowd, Skelton et al., 2012). This delay has significant consequences for investigative interviews and the accuracy of information recalled by witnesses. The longer the delay, the greater the risk of memory decay and forgetting (Ellis et al., 1980; Murre & Dros, 2015). As time passes and memories decay, witnesses struggle to recall or recognise facial details (Frowd et al., 2015; Hine & Itoh, 2018; Krouse, 1981; Shepherd & Ellis, 1973; Shepherd et al., 1991). In the context of this thesis, inaccurate memories can have serious consequences for criminal procedures. Inaccurate witness accounts can lead to wrongful arrests and potentially imprisonment of innocent individuals, whilst the true offenders have opportunity to continue offending (e.g., Innocence Project, 2023).

Many techniques have been developed to aid memory retrieval in witnesses, particularly when dealing with long retention intervals (e.g., the CI and the H-CI). However, memory decay can still be detected even when these facilitating techniques are applied after 24 hours (Frowd, Carson, Ness, McQuiston-Surrett et al., 2005). Supporting this is a meta-analysis conducted by Frowd et al. (2015), comprised of 23 studies, which demonstrated that composites constructed using the CI or H-CI were less accurate when the retention interval was lengthy (e.g., 1-2 days) compared to shorter delays (e.g., 0-3.5).

As elaborated, the retention interval heavily impacts memory retention and retrieval processes. It is often queried about how long the retention interval should be to achieve optimum accuracy. Brown et al. (2020) investigated the impact of both short and long retention intervals on memory accuracy. Participants in their experiment observed an unfamiliar face, and then provided a verbal description of the facial details either 3-4 hours later, or after a 2-day delay. Subsequently, participants either constructed a composite instantaneously, or after a 30-minute delay. Results revealed that participants who provided a verbal description 3-4 hours after seeing the unfamiliar face, and constructed the composite immediately, constructed more accurate composites than those subjected to a longer retention interval of 2 days (between seeing the unfamiliar face, and providing a description of facial features) and a delay of 30 minutes (between providing the facial description and constructing the composite).This research supports the notion that, as time passes memory can decay, which opens doors for error.

Frowd, Carson, Ness, McQuiston-Surrett et al. (2005) conducted an experiment to investigate the impact of a typical forensic retention interval (two days), on subsequent composite accuracy. Participants observed an unfamiliar face for 60 seconds and then, two days later, constructed a composite of the face. They proposed that, as the composite construction process involves the difficult task of recall (e.g., providing a verbal description of the face) and recognition (e.g., recognising the facial features when constructing a composite), employing a longer retention interval would likely result in less accurate composites to those constructed with a shorter delay (e.g., 20% identification rate after 3-4 hours; Frowd, Carson, Ness, Richardson et al., 2005). Findings supported this hypothesis and revealed that the overall percentage of correctly identified composites was very low at 3% with the typical forensic delay of 48 hours. Although a small likeness to the target face could be accomplished, the low identification rates could lead to practical complications, including misidentifications, with long-term detrimental effects on individuals (e.g., Innocence Project, 2023).

The previous two experiments in this thesis employed either the CI or H-CI with a 24-hour retention interval, potentially contributing to the lower-than-expected naming accuracy scores (cf., Frowd et al., 2013; Frowd et al., 2019). The foregoing information forms the foundation for the current experiment, proposing that implementing a shorter retention interval, alongside the memory facilitating technique of the H-CI, may enhance participants memory and, consequently, the accuracy of the facial composites they construct.

Multiple retrieval attempts and timing

Contrary to the foregoing research, it has been discovered that longer intervals between retrieval attempts can, under the right circumstances, often produce accurate information, with little errors (Roediger et al., 2009). One important factor that can alleviate the negative impacts of extended delays is the application of repeated retrieval attempts (Murre & Dros, 2015). When witnesses are given multiple opportunities to retrieve information, it greatly supports witnesses in accessing memories that may have been initially forgotten (Odinot & Wolters, 2006; Roediger et al., 2009).

For some time now, psychologists have investigated how the timing and spacing of repeated retrieval opportunities impact the long-term retention of memories, specifically, the timing that is most effective (Hintzman & Rogers, 1973). The spacing effect suggests that repeated retrieval attempts, spaced out over a lengthier time, tend to enhance retrieval accuracy, whereas closely spaced attempts may decrease accuracy later on (Roediger et al., 2009).

Butler and Roediger (2007) conducted an experiment investigating the impact of multiple retrieval opportunities on long-term memory retention. Within this experiment, participants attended three lectures (on consecutive days), and engaged in different activities after each lecture, including studying a lecture summary, or taking either a multiple-choice or a short answer test. A concluding short answer test was carried out one month later. Results showed that taking multiple-choice tests in close proximity to encoding information improved memory retrieval more so than not practising. However, those participants who engaged in short answer tests after the lectures, retrieved the most accurate information. This finding is a great illustration of how multiple retrieval attempts, in short proximity to learning, can actually advance retrieval of information over a period of time.

In a related study, Roediger and Karpicke (2006) conducted two experiments where students learned information passages and either took one or three free recall tests immediately after learning, or simply restudied the information the same number of times as the tested group. Students then took a final recall test either after 5 minutes, 2 days, or a week after the initial learning. Memory retrieval was most improved when the test was administered after a 5-minute retention interval, compared to the longer delays. On the delayed tests, multiple retrieval opportunities were more effective than repeated studying. While repeated studying had a positive impact on students' confidence in their recall abilities, repeated testing emerged as the more powerful technique for memory consolidation.

Although it has been suggested that, to obtain the most optimum results from memory, retrieval attempts should be spaced apart, (Hintzman & Rogers, 1973; Roediger et al., 2009), the foregoing studies suggest otherwise. Specifically, they elaborate that if the initial retrieval attempt closely follows the encoding of information, it can still enhance recall, even after a lengthier period of time. These findings may be explained by the concept of the forgetting curve, initially developed by psychologist Ebbinghaus (Murre & Dros, 2015). The forgetting curve describes how memory recall tends to decline over time if individuals make no conscious effort to reinforce the information (Murre & Dros, 2015). The biggest deterioration in the forgetting curve is most significant, shortly after the initial encoding of information. Murre and Dros (2015) replicated Ebbinghaus's findings, reinforcing the validity of the forgetting curves concept. A key insight regarding the forgetting curve is that the degree of forgetting can be relatively reduced by giving witnesses multiple opportunities to practice and retrieve information. According to Ebbinghaus, the initial of these retrieval sessions should occur when recall has slowly started to fade, but not enough to where information has been forgotten (Murre & Dros, 2015).

Current research

The literature reveals that both the retention interval and multiple retrieval attempts can impact the quality and quantity of information that a witness can retrieve. However, it is commonly recommended that shorter-retention intervals are the most influential on accurate memory recall (e.g., Brown et al., 2020; Ellis et al., 1980; Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Frowd, Carson, Ness, Richardson et al., 2005; Frowd et al., 2015; Hine & Itoh, 2018; Krouse, 1981; Shepherd & Ellis, 1973; Shepherd et al., 1991). Though, it is important to note that under the right circumstances (Hintzman & Rogers, 1973; Murre & Dro's, 2015; Roediger et al., 2009) longer retention intervals can be advantageous when combined with multiple retrieval attempts at strategic intervals.

The current experiment will employ a shorter retention interval of 3-4 hours. In addition, participants will be given multiple opportunities to make RBE to the internal facial features of the composite. This approach is proposed with the expectation that combining a short retention interval and multiple opportunities to make RBE, will enhance the overall accuracy of composites. Specifically, it is hypothesised that the more opportunity that participants are given to make RBE to the composite (i.e., A>B>C>D), the more accurate the composite will be, as compared to those composites constructed with limited opportunity to do so.

METHOD

As with the previous three experiments, Experiment 4 was divided into two Stages. The first Stage involved participants constructing a composite of an unfamiliar face. Stage 2 involved a naming task, and a rating task, to evaluate the accuracy of the composites constructed in Stage 1.

Ethics

Experiment 4 received ethical approval from the ethics board within the School of Psychology at UCLan.

Design

The design was the same as in the previous three experiments (see *Design* section in Chapter 2 for more detail).

STAGE 1: COMPOSITE CONSTRUCTION

Participants

Thirty participants, 18 females and 12 males (*M* = 32.3, *SD* = 13.0), who were unfamiliar with the television soap Coronation Street were recruited, in the same way as the previous experiments, via opportunity sampling. Participants who were recruited from SONA received 7 SONA points for their participation, whilst other participants received a £5 Amazon voucher.

Materials

The target-photographs were selected in the same way as the previous experiments. However, this time the target-identities were actors and actresses from the television soap Coronation Street (as in Experiment 1 and 2). The researcher previously had difficulty recruiting participants who were familiar with EastEnders, and found Coronation Street easier for familiarity recruitment. The target identities were Peter Barlow, Tracey Barlow, Leanne Battersby, Fiz Brown, Roy Cropper, Eileen Grimshaw, Steve McDonald, Sally Metcalfe, David Platt, and Kirk Sutherland. The target-photographs were 8cm x 10 cm and in colour.

Procedure

The experiment was conducted in the same way as Experiment 3, with the H-CI. However, one key difference was that the retention interval was reduced from 24 hours to 3-4 hours. Instead of inviting participants back the following day, they constructed the composite on the same day. Apart from this slight modification, the procedure for the current experiment followed that of Experiment 3. On the day of the experiment, participants were invited to look at one of the ten target-identities that they were unfamiliar with for a total of 60 seconds. Three to four hours later, the participant was invited back to construct a composite. Before constructing a composite, a H-CI was administered, following the same procedure as Experiment 3. Composites were also constructed in the same way as the previous experiments, using the same four conditions (i.e., **AEFS**).

STAGE 2: COMPOSITE EVALUATION

NAMING TASK

The naming task involved participants looking at an array of composites and attempting to identify them.

Participants

Twenty-eight participants, 21 females and seven males (*M* = 42.1, *SD* = 16.3), who were familiar with Coronation Street were recruited in the same way as previous experiments, via opportunity sampling. Participants who were recruited through SONA received 1.5 SONA points for their participation, whereas participants recruited through recruitment websites received a £2 Amazon voucher.

Materials and Procedure

The materials and procedure for the naming task were the same as in previous experiments. As part of the naming task, an *a priori* rule was specified. The rule specified that participants needed to recognise 80% of the ten target-photographs. All participants met this specification and therefore all data were included in the analysis. There was also no missing data.

LIKENESS TASK

The likeness task involved participants rating the accuracy of composites against the target face on a Likert scale of 1 (low likeness) to 7 (high likeness).

Participants

Fifty-four participants, 26 females and 28 males (*M* = 25.0, *SD* = 7.6) who were familiar with the television soap Coronation Street were recruited via opportunity sampling. Participants who were recruited through SONA received 1.5 SONA points for their time, whereas those recruited through recruitment websites received a £2 Amazon voucher.

Materials and Procedure

The materials and procedure for the likeness task were the same as in Experiment 3. Composites were analysed for the whole face (i.e., internal, and external features), internal features (e.g., eyes, nose, and mouth), and external features (e.g., ears, face shape and hair).

RESULTS

ANALYSIS

The traditional ANOVA is adopted throughout the thesis to analyse the data. However, the traditional ANOVA does not provide the evidence needed for both the null and alternative hypothesis; therefore, Bayesian Factors ANOVA will be adopted throughout the thesis in order to provide that information.

MANIPULATION CHECK

Target-photographs were named well (*M* = 100.0, *SD* = 0.0) with no difference between **AEFS** meaning that all participants had the same potential to correctly identify all composites in each **AEFS**. The naming scores are summarised in Table 20.

Table 20: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Target-Photograph Naming.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
100.0	100.0	100.0	100.0
(0.0)	(0.0)	(0.0)	(0.0)

As a manipulation check, correct target-photograph naming scores by **AEFS** were analysed using an independent samples ANOVA. All participants correctly named each of the target-photograph within each **AEFS**. This suggests that there was no significant difference between correct target-photograph naming scores by **AEFS**.

NAMING TASK

Responses were scored for accuracy. A score of 1 was given for correct responses and 0 if no name, or an incorrect name was provided. Participants responses were checked for missing data, of which none was found. Overall, correct composite naming scores were 26.8% (*SD* = 14.2). The naming data were analysed both by participants and by items (see *Naming Task* section in Chapter 2 for description). It was hypothesised, in agreement with existing research (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984) that providing participants with more opportunity to make RBE (i.e., A>B>C>D) to the facial composite would increase overall composite quality, and, further, the greater the percentage of correct composite naming scores would be.

By participants analysis

Correct composite naming percentages were analysed by participants. The percentage of correctly named composites, by **AEFS**, was analysed using an independent samples ANOVA. An independent samples Bayesian Factors ANOVA was also conducted to find the BF value. The naming scores are summarised in Table 21.

Table 21: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Correct Composite Naming: By Participants.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
30.0	20.0	20.0	37.1
(10.0)	(14.1)	(14.5)	(15.0)

There was a non-significant main effect on correct composite naming scores by **AEFS** [$F(3,24) = 3.00$, $p = .052$, BF = 0.40], suggesting that there was no significant difference between correct composite naming scores and **AEFS**. The BF of 0.40 also reveals anecdotal evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their correct naming scores.

By items analysis

Correct composite naming percentages were analysed by participants. The percentage of correctly named composites, by **AEFS**, was analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The naming scores are summarised in Table 22.

Table 22: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Naming Scores: By Items.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
2.1	1.4	1.4	2.6
(1.6)	(0.8)	(1.1)	(1.6)

There was a non-significant main effect on correct composite naming scores by **AEFS** [$F(3,27) = 2.10$, $p = .168$, BF = 0.15], suggesting that there was no significant difference between correct composite naming scores and **AEFS**. The BF of 0.15 also reveals moderate evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their correct naming scores.

LIKENESS TASK

Participants rated the composites, on a Likert scale, of 1 (low likeness) to 7 (high likeness) for like likeness to the target-photograph.

As with Experiment 3, likeness scores were analysed for the constructed composites; whole face (both internal and external features), internal features (e.g., eyes, nose, and mouth) and external features (e.g., ears, face shape and hair). Analysis was also conducted both by participants and by items (see *Likeness Task* in Chapter 2 for description). It was hypothesised, in agreement with existing research (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984) that providing participants with more opportunity to make RBE (i.e., A>B>C>D) to the facial composite would increase overall composite quality, and, further, the greater the likeness scores would be.

WHOLE FACE ANALYSIS

By participants analysis

Likeness scores were analysed, by participants, for the whole face of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 23.

Table 23: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Participants for Whole Face.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.9	3.6	3.9	4.2
(0.8)	(0.7)	(0.7)	(0.6)

Note: The likeness scale ranged from 1 (low likeness) to 7 (high likeness). For the by participants analysis one group difference was significant **AEFS B** and **AEFS D** (p = .001).

There was a significant main effect on likeness scores by **AEFS** [$F(3,51) = 3.90$, p =.014, BF = 1.90]. The BF of 1.90 also reveals anecdotal evidence (e.g., Lee & Wagenmakers, 2014) for the alternative hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did influence the accuracy of composites and the overall likeness scores.

Follow up tests were conducted with a Bonferroni Correction ($\alpha < .05/6$). One of the six paired-samples t-tests (two-tailed) revealed a significant difference between composite likeness scores and **AEFS.** This significance was between **AEFS B** and **AEFS D** ($p = .001$). All remaining t-tests were non-significant ($p > .075$). In more detail, **AEFS D** produced significantly more accurate composites than **AEFS B** (with a large ES = 0.91), **s**uggesting that the composites that were constructed with a greater opportunity to make RBE, produced more accurate composites than those created with less opportunity. This supported the experimental hypothesis.

By items analysis

Likeness scores were analysed by items for the whole face of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 24.

Table 24: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Items for Whole Face.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.9 (0.7)	3.6 (0.8)	3.9 (0.5)	4.2 (0.5)

There was a non-significant main effect on likeness ratings by **AEFS** [$F(2.03,18.29) = 2.02$, $p = .161$, BF = 0.27], suggesting that there was no significant difference between likeness scores and **AEFS**. The BF of 0.27 also reveals moderate evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their likeness scores.

INTERNAL FEATURES ANALYSIS

By participants analysis

Likeness scores were analysed, by participants, for the internal facial features of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 25.

Table 25: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Participants for Internal Features.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.9	3.7	4.1	4.4
(0.7)	(0.6)	(0.6)	(0.6)

Note: The likeness scale ranged from 1 (low likeness) to 7 (high likeness). For the by participants analysis three group differences were significant **AEFS A** and **AEFS D** (p = .001), **AEFS B** and **AEFS C** (p = .004), **AEFS B** and **AEFS D** (p <.001).

There was a significant main effect on likeness scores by **AEFS** [$F(3,51) = 11.69$, p <.001, BF >100]. The BF of >100 also reveals extreme evidence (e.g., Lee & Wagenmakers, 2014) for the alternative hypothesis Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did influence the accuracy of composites and their likeness scores.

Follow up tests were conducted with a Bonferroni Correction ($\alpha < .05/6$). Three of the six paired-samples t-tests (two-tailed) revealed a significant difference between composite likeness scores and **AEFS** (**AEFS A** and **AEFS D**, p = .001; **AEFS B** and **AEFS C**, p = .004, **AEFS B** and **AEFS D**, p <.001). All remaining t-tests were non-significant (p > .05). In more detail **AEFS D** was significantly greater than **AEFS A** (with a large ES = 0.91) and **AEFS B** (with a large ES = 1.4). Further, **AEFS C** was significantly greater than **AEFS B** (with a medium ES = 0.78). This suggests that the composites that were constructed with a greater opportunity to make RBE produced more accurate composites than those created with less opportunity. This supported the experimental hypothesis.

By items analysis

Likeness scores were analysed, by items, for the internal facial features of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 26.

Table 26: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Items for Internal Features.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.9	3.7	4.1	4.4
(0.5)	(0.7)	(0.4)	(0.5)

Note: The likeness scale ranged from 1 (low likeness) to 7 (high likeness). One group difference was significant **AEFS B** and **AEFS D** (p = .003).

There was a significant main effect on likeness scores by **AEFS** [$F(3,27) = 5.71$, $p = .004$, BF = 36.10]. The BF of 36.10 also reveals very strong evidence (e.g., Lee & Wagenmakers, 2014) for the alternative hypothesis. In regard to the hypothesis, this shows that providing participants with more opportunity to make RBE to internal facial features of a composite did influence the accuracy of composites and likeness scores.

Follow up tests were conducted with a Bonferroni Correction ($\alpha < .05/6$). One of the six paired-samples t-tests (two-tailed) revealed a significant difference between composite likeness scores and **AEFS** (**AEFS B** and **AEFS D**, $p = .003$). All remaining t-tests were non-significant ($p > .019$). In more detail, **AEFS D** was significantly greater than **AEFS B** (with a large ES = 1.3). This suggests that the composites that were constructed with a greater opportunity to make RBE produced a more accurate composite than those created with less opportunity. This supported the hypothesis.

EXTERNAL FEATURES ANALYSIS

By participants analysis

Likeness scores were analysed, by participants, for the external facial features of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 27.

Table 27: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Participants for External Features.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.9 (0.9)	4.1 (0.6)	4.4 (0.7)	4.2 (0.7)

There was a non-significant main effect on likeness ratings by **AEFS** [$F(3,51) = 2.67$, $p = .057$, BF = 0.40], suggesting that there was no significant difference between likeness scores and **AEFS**. The BF of 0.40 also reveals anecdotal evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this revealed that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their likeness scores.

By items analysis

Likeness scores were analysed, by items, for the external facial features of a composite. The likeness scores of the composites, by **AEFS**, were analysed using a repeated measures ANOVA. A repeated measures Bayesian Factors ANOVA was also conducted to find the BF value. The likeness scores are summarised in Table 28.

Table 28: Mean and (Standard Deviation) for Stage of Construction (**AEFS**) on Composite Likeness Scores: By Items for External Features.

Stage of Construction			
AEFS A	AEFS B	AEFS C	AEFS D
3.9 (0.9)	4.1 (0.8)	4.4 (0.7)	4.2 (0.8)

There was a non-significant main effect on likeness ratings by **AEFS** [$F(1.83,16.44) = 0.19$, $p = .189$, BF = 0.40], suggesting that there was no significant difference between likeness scores and **AEFS**. The BF of 0.40 also reveals anecdotal evidence (e.g., Lee & Wagenmakers, 2014) for the null hypothesis. Regarding the hypothesis, this reveals that providing participants with more opportunity to make RBE to the internal facial features of a composite did not influence the overall accuracy of the composites, or their likeness scores.

DISCUSSION

Research has shown that the length of time between information retention and retrieval, specifically when combined with multiple retrieval attempts, can have an advantageous influence on the quality and quantity of information recalled by a witness (e.g., Brown et al., 2020; Butler & Roediger, 2007; Ellis et al., 1980; Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Frowd, Carson, Ness, Richardson et al., 2005; Frowd et al., 2015; Hine & Itoh, 2018; Murre & Dros, 2015; Odinot & Wolters, 2006; Roediger & Karpicke, 2006; Roediger et al., 2009). It was proposed that reducing the retention interval to a shorter delay of 3-4 hours might prove beneficial. It was predicted that this, combined with the H-CI and multiple opportunities to provide RBE, would enhance the overall accuracy of the resulting composites. The findings from the current experiment partially supported this hypothesis. The initial hypothesis proposed that providing participants with more opportunity to make RBE to the internal facial features of a composite would increase the overall percentage of correctly named composites. However, the results did not support the hypothesis.

The second hypothesis proposed that providing participants with more opportunity to make RBE to the internal facial features of a composite would increase the overall likeness scores of the resulting composite. This hypothesis received support. First, when analysing the whole face, it was identified that in the by participants analysis, there was a reliable difference between **AEFS B** and **AEFS D**. Specifically, **AEFS D** produced more accurate composites than those constructed in **AEFS B**. This suggests that providing participants with more opportunity to make RBE to the internal facial features of a composite did improve the overall accuracy of the composites.

Second, when examining the analysis of the internal facial features, it was discovered in the by participant analysis, **AEFS D** produced more accurate composites than **AEFS A** and **AEFS B**. Moreover, **AEFS C** produced more accurate composites than **AEFS B**. In regard to the by items analysis, **AEFS D** produced more accurate composites than **AEFS B**. These findings further support those from the whole face analysis, and the experimental hypothesis, that providing participants with more opportunity to provide RBE to the internal facial features would lead to more accurate composites.

Importantly, for the first time throughout this thesis, the conditions employed in the current experiment, which included the H-CI, multiple opportunities to make RBE, and a short retention interval, led to **AEFS C** being one of the most accurate Stages. These findings are particularly important for forensic practice, as they highlight the optimal conditions that should be applied during investigative interviews and composite construction sessions to achieve the best results, specifically regarding the stage where the composite is typically saved and distributed from.

The current experiment's findings align with existing research in various ways. The findings confirm that implementing a short retention interval, alongside providing participants with more opportunity to retrieve information (specifically, by providing them with more opportunity to make RBE), does contribute to improved witness memory (e.g., Butler & Roediger, 2007; Murre & Dros, 2015; Odinot & Wolters, 2006; Roediger & Karpicke, 2006; Roediger et al., 2009), and further, the accuracy of constructed composites (e.g., Brown et al., 2020; Frowd et al., 2015).

These findings support the concept of the forgetting curve (e.g., Murre & Dros, 2015). The forgetting curve posits that in order for witnesses to provide an accurate and reliable account of information, their initial recall attempt (e.g., during an investigative interview) must be in close proximity to when the information was first encoded (e.g., at the crime scene). If information is not rehearsed before the curve of forgetting begins to decline, it will be forgotten (e.g., Ebbinghaus in Murre & Dros, 2015). Without rehearsal, the information may not be remembered at subsequent retrieval attempts. This could explain the findings from the previous experiments where a 24-hour retention interval was employed, leading to the potential loss of valuable information about the target face (e.g., Frowd, Carson, Ness, McQuiston-Surrett et al., 2005). It is possible that the task of providing participants with more opportunities to make RBE was difficult as participants potentially did not remember enough detail, initially, to be strengthened by the multiple opportunities to make RBE. Regarding the current experiment, the shorter retention interval may have permitted participants to retrieve information before the forgetting curve began to decline. Consequently, as the composite software progressed, coinciding with the opportunity to provide RBE, participants could retrieve more information about the target face, resulting in more accurate composites.

Psychologists, have, for some time, studied the timing and spacing of repeated retrieval opportunities. A longer retention interval between encoding and retrieval can negatively impact face recognition at a later point (Murre & Dros, 2015). This factor may account for inconsistencies throughout the **AEFS**, with **AEFS A** through the previous three experiments producing the most accurate composites. It emphasises that while multiple retrieval opportunities are crucial for witness memory, the timing and spacing of these opportunities must be carefully considered as supported within the findings throughout this thesis. These findings support the findings in the meta-analysis by Frowd et al. (2015) who found that memory enhancing techniques, such as the CI and H-CI are most effective when combined with a shorter retention interval.

Limitations and future directions

The percentage of correct composite naming scores from the current experiment have returned to similar levels as seen in Experiment 2, which were the highest levels observed throughout this thesis. This suggests that the implementation of the H-CI, alongside the shorter retrieval interval, did contribute to an increase in the percentage of correct composite naming scores when given more opportunity to make RBE. However, these results are still far from existing research. It is important to note however, that when comparing all the experiments conducted in this thesis to the percentages of correct composite naming statistics in current research using EvoFIT (e.g., Frowd et al., 2013) there are some contextual differences to consider. These experiments were conducted face-to-face, in a controlled environment (e.g., Elliot et al., 2022) where factors such as attentiveness and distraction could be controlled and monitored effectively. Considering this, it would be valuable for future research to replicate the current experiment but within a controlled laboratory setting. This would help to determine if conducting the experiment under controlled conditions would bring the composite identification numbers closer to those observed in research. It would be important for the researchers to assess composites with a higher level of accuracy, so that each **AEFS** can be looked at more reliably.

Furthermore, when reviewing the literature for Experiment 4, it becomes evident that providing witnesses with multiple retrieval opportunities, spread over a period of time, can enhance memory recall, by strengthening memories (Butler & Roediger, 2007; Murre & Dros, 2015; Roediger & Karpicke, 2006). Another promising avenue for future research could involve spreading retrieval opportunities over a lengthier period of time, rather than conducting them all in a single interviewing session. For instance, coinciding with the forgetting curve (e.g., Murre & Dros, 2015) researchers could test participants within a close proximity to encoding a target face and then at different retrieval intervals, before constructing a composite of the face. It could help to determine whether a combination of immediate recall and subsequent retrieval opportunities (through a CI or H-CI), and providing witnesses with more opportunity to make RBE (during composite construction) could collectively improve the overall accuracy of the resulting composites.

Witnesses who are expected to participate in multiple interviews over an extended period of time (often seen in legal contexts such as court proceedings) could benefit considerably from the implementation of shorter retention intervals. However, while there is a substantial body of evidence supporting the effectiveness of a shorter retention interval their practical application may be obstructed by real-world constraints. The primary challenge lies in the limited availability of police resources (e.g., Fodarella et al., 2015; Frowd, Pitchford et al., 2012; Frowd, Skelton et al., 2012). This practical implication raises concerns about the viability of widespread implementation. To address this issue, and provide a stronger argument for the need of refining criminal procedures to improve witness memory accuracy, a potential avenue is to replicate the experiment in a controlled laboratory environment. If the results dependably align with those found in existing research, and this thesis, it could provide stronger support for the need to review, and potentially revise, criminal procedures. Ultimately, this would contribute to the predominant purpose of enhancing the accuracy of witness memory in criminal procedures.

Experiment 4 had the objective of determining whether providing participants with more opportunity to make RBE to the internal facial features of a composite, with the use of a H-CI and the implementation of a shorter retention interval, would lead to a subsequent improvement in the overall accuracy of the composite. Overall, it is evident that using memory facilitating techniques, alongside a short retention interval, whilst, providing participants with multiple opportunities to make RBE, leads to an overall increase in composite accuracy, specifically in regard to the Stage that composites are usually saved and distributed from (**AEFS C**). From a practical standpoint, these results suggest that it would be advantageous for witnesses to be interviewed shortly after witnessing a crime, with an initial attempt to recall details as this approach appears to enhance the accuracy of the composites constructed.

CHAPTER 6: GENERAL DISCUSSION

The primary focus of this thesis was to investigate innovative methods with the potential to enhance witness memory during investigative interviews and composite construction procedures. The central hypothesis investigated was whether increasing the opportunity that witnesses had to make RBE to the internal facial features of a composite would lead to more accurate composites. This hypothesis was grounded in extensive pre-existing research (e.g., Brown et al., 2017; Butler & Roediger, 2007; Fisher & Geiselman, 1992; Geiselman et al., 1984; Roediger & Karpicke, 2006; see Köhnken et al., 1999; Memon, Meissner et al., 2010 for a meta-analysis) surrounding the principle that providing witnesses with multiple opportunities to retrieve information can effectively enhance witness memory and information recall (see foregoing Chapters for more detail). While this approach had been studied in the context of investigative interviewing (e.g., the CI), its application to the composite construction process had not yet been explored. Drawing insights from existing research, the researcher proposed that it would be valuable to explore this avenue, to identify whether it would impact the accuracy of composites.

To accomplish this, a comprehensive analysis of each component of the EvoFIT software was conducted to determine whether increasing opportunities to make RBE to the internal facial features of the composite increased overall composite accuracy. Additionally, this thesis aimed to examine whether incorporating memory facilitating techniques, commonly used in practice (e.g., the CI: Fisher & Geiselman, 1992; Geiselman et al., 1984; the H-CI: Frowd, Bruce, Smith et al., 2008; short retention interval: Murre & Dros, 2015., Frowd, Bruce, Ness et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Frowd, Carson, Ness , Richardson et al., 2005; Frowd et al., 2015), would further influence composite accuracy. Ensuring the composite software is producing highly accurate composites is vital for effective offender identification and reducing the risk of misidentification (e.g., Innocence Project, 2023).

Throughout this thesis, four experiments were methodically designed to rigorously test the hypothesis; that providing witnesses with more opportunity to make RBE to the internal facial features of a composite could improve overall composite accuracy. Although each experiment aligned with this core hypothesis, slight variations were introduced to replicate common techniques used in composite construction processes within policing practice. The following sections will delve into the detailed findings of each of those experiments.

Experiment 1

The main objective of Experiment 1 was to test the experimental hypothesis using the WAH version of EvoFIT (see Martin et al., 2018 for a review). The adoption of the WAH version was driven by the constraints imposed by the COVID-19 pandemic, which made face-to-face experiments unfeasible at the time. This adaption permitted research to continue and enabled the remote construction of facial composites.

While the use of the WAH software offers many potential benefits (e.g., Martin et al., 2018) it also introduces challenges, by placing the responsibility of composite construction solely on the participants, with no researcher involvement. This shift in responsibility had a notable limitation in that memory enhancing techniques typically employed in investigative interviews (e.g., the CI and H-CI) could not be implemented prior to the composite construction process, due to the absence of the researcher. Although the experiment allowed researchers to investigate whether providing witnesses with more opportunity (an important component of the CI) to make RBE would increase composite accuracy, it prevented other components of the CI from being used. For instance, the absence of rapport building, an important factor known to enhance witness memory (e.g., Fisher & Geiselman, 1992) by making the witness feel at ease (Fodarella et al., 2015; Vallano & Schreiber Compo, 2011), reducing levels of anxiety (Kieckhaefer et al., 2014) and minimising cognitive load by allowing the witness to focus on the task rather than potential anxieties about the procedure (Eysenck, 1979; Eysenck & Calvo, 1992; Nahouli et al., 2021) which may potentially explain why the percentages of correctly named composites were low in this experiment.

However, the absence of memory facilitating techniques did enable the researchers to evaluate the effectiveness of each **AEFS** in isolation. Surprisingly, the outcomes of Experiment 1 contradicted the initial hypothesis, and diverged from a substantial body of existing research in this field (e.g., Brown et al., 2017; Butler & Roediger, 2007; Fisher & Geiselman, 1992; Geiselman et al., 1984; Roediger & Karpicke, 2006; see Köhnken et al., 1999; Memon, Meissner et al., 2010 for a meta-analysis; see foregoing Chapters for more detail). In terms of the findings, it became evident that the Stage where the composite is typically saved and distributed (**AEFS C**) did not produce the most accurate composites and, contrary to expectations, composites constructed with minimal opportunity to make RBE to the internal facial features were more accurate. In light of the findings, the results suggests that the WAH process could be refined to allow witnesses to construct more accurate composites without having to navigate the entire procedure.

Furthermore, the introduction of **AEFS D**, aimed at providing participants with more opportunity (than the software usually permits) to make RBE to the internal facial features, produced composites with reliably higher likeness scores than **AEFS C**. Incorporating **AEFS D** into the typical EvoFIT WAH procedure could enhance composite accuracy by offering witnesses more opportunities to make enhancements to the composite until likeness is achieved. Although this finding did not entirely align with the initial hypothesis, these findings did support past research advocating for the implementation of multiple retrieval opportunities to enhance memory accuracy.

Despite not supporting the proposed hypothesis, the findings do carry importance. The findings highlight that there are distinct variations in composite accuracy across the different Stages of the EvoFIT process. Composite accuracy fluctuated throughout the procedure, challenging the assumption that accuracy consistently improves with opportunity to make RBE, and that the point of typical composite distribution (**AEFS C**) does not produce the most accurate composites. This may explain the misidentifications that are seen in practice (e.g., Innocence Project, 2023). This insight suggests the prospect of replicating the experiment to validate the results that were found and to explore the potential feasibility of modifying the composite construction procedure. Specifically, when using the WAH system, the procedure could be shortened without the risk of a degree in accuracy.

This observation laid the groundwork for Experiment 2, which proposed that the researcher should be actively involved in the composite construction process. The involvement permitted the researcher to use the memory facilitating technique of the CI (e.g., Fisher & Geiselman, 1992), a standard technique used in investigative interviews, thereby opening the door for the incorporation of other crucial components. For instance, by implementing the CI, it becomes possible to attempt to establish rapport between the researcher and participant. One key advantage of building rapport is the potential reduction that it has on anxiety levels during the interaction. Lower anxiety levels can contribute to a decrease in cognitive load which in turn can enhance participants memory retrieval (Carter et al., 1996; Eysenck, 1979; Eysenck & Calvo, 1992; Kieckhaefer et al., 2014) and hopefully permit participants to retrieve more information at each opportunity to make RBE.

Experiment 2

Experiment 2 had two main objectives: first, to advance deeper into the experimental hypothesis laid out throughout the thesis, and second, to build upon the insights gained from Experiment 1. The primary focus of Experiment 2 was to improve the percentage of correctly named composites, so that the findings from the analyses could be taken with more confidence. It is again worth noting that Experiment 2 took place during ongoing COVID-19 restrictions, which limited in-person research. In response to the low correct naming scores observed in Experiment 1, the researchers decided to refine the methodology for Experiment 2 by incorporating memory enhancing techniques commonly used in witness memory research and practice, aiming to align the results more closely with established research findings from similar studies (e.g., Frowd et al., 2013) and practice (e.g., Frowd et al., 2019).

In light of these considerations, one meaningful change in Experiment 2 was the active involvement of the researcher during the procedure, allowing for the implementation of memory enhancing techniques. This shift involved changing from the previously used WAH procedure (e.g., Martin et al., 2018), which was participant-led, to a practitioner-led approach using the screen sharing tool on Microsoft Teams. Through screen sharing, the researcher could access and control the software, allowing for a more active role. This approach also permitted the researcher to apply the CI technique, a well-established method for enhancing memory recall (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984).

Interestingly, despite not fully supporting the initial experimental hypothesis, a trend parallel to Experiment 1 emerged. Providing participants with more opportunity to make RBE to the internal facial features of the composites did not lead to an overall enhancement in composite accuracy. Surprisingly, composites constructed from **AEFS A** and **AEFS B** exhibited a greater likeness towards the target face compared to those from **AEFS C**, where the composite is typically saved and distributed from. Intriguingly, another similar pattern to Experiment 1 emerged, with **AEFS D** producing more accurate composites than those constructed in **AEFS C**. This finding emphasises that, while an increased opportunity to make RBE to the internal facial features of a composite does contribute to constructing more accurate composites, this does not necessarily translate into producing the most accurate composite. These trends suggest that adjustments could be made to the practitioner-led version of the composite construction process when combined with the CI technique, potentially eliminating the need for witnesses to engage with the whole software to construct an accurate composite.

Although the original hypothesis was not fully supported, there was a notable improvement in correct composite naming percentages, which increased to 27.0%, compared to Experiment 1 findings of 16.6%. This improvement highlights the effectiveness of implementing a practitioner-led method combined with the CI technique to enhance witness memory, aligning with existing research on memory enhancement techniques (e.g., Fisher & Geiselman, 1992; Geiselman et al., 1984). While these advancements have not yet begin to rise to the levels seen in existing research (e.g., Frowd et al., 2013), they do lend greater credibility to the findings.

Despite not fully meeting the proposed hypothesis, these findings remain valuable and informative. The outcomes continue to highlight the distinct disparities present across the different Stages of the EvoFIT process and suggest that composite accuracy does fluctuate throughout the construction procedure, rather than produce a linear increase as RBE opportunities increase. These insights suggest that future replications of the experiment may be necessary to validate these results and assess the practicality of adapting the composite construction procedure when using a practitioner-led method alongside the CI technique. This approach could potentially eliminate the need for witnesses to engage with the whole system during composite construction.

In the broader context of this thesis, these findings influenced the direction for Experiment 3. Given that the percentage of correctly named composites was still relatively low, the researchers proposed integrating a different memory-enhancing technique commonly used in research, known as the H-CI (e.g., Frowd, Bruce, Smith et al., 2008).

Experiment 3

In Experiment 3, the aim was to replicate the methods used in Experiment 2, whilst introducing an alternative memory facilitating technique known as the H-CI (e.g., Frowd, Bruce, Smith et al., 2008). This decision was influenced by insights drawn from the broader literature (discussed throughout Chapter 1 and expanded upon through Chapter 4). The literature highlighted the advantages of the H-CI, particularly its alignment with facial recognition systems and its ability to facilitate the retrieval of accurate information and produce accurate composites when compared to using the CI alone (Frowd, Bruce & Hancock, 2008; Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd et al., 2004; Frowd, McQuiston-Surrett et al., 2005; Frowd, Nelson et al., 2012).

The findings from Experiment 2 indicated that combining the researcher's involvement with the memory enhancing technique of the CI led to an increase in composite accuracy, as demonstrated by a higher percentage of correctly named composites. Building on this, the H-CI was introduced with the expectation that it would outperform the CI and further enhance the accuracy of correctly named composites. Additionally, it was essential to assess the impact of the H-CI on the experimental hypothesis of this thesis, as it is frequently used in composite construction research. Researchers wanted to determine whether its application influences the individualised Stages of EvoFIT, and identify the point where optimal composite accuracy was achieved. However, Experiment 3 produced unexpected results that contradicted the proposed hypothesis. Contrary to expectations, based on existing literature which suggested that the H-CI would outperform the CI (Frowd, Bruce & Hancock, 2008; Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd et al., 2004; Frowd, McQuiston-Surrett et al., 2005; Frowd, Nelson et al., 2012) the combination of the H-CI, with an increased opportunity to make RBE to the internal facial features of the composite, did not lead to higher percentages of correctly named composites (23.9%) in comparison to Experiment 2, which solely employed the CI.

In the preceding two experiments, no consistent differences were found throughout the **AEFS** for the naming task. However, distinctions emerged throughout the **AEFS** during the likeness tasks. This prompted researchers to restructure the likeness task analysis for the current experiment, aiming to isolate features, and identify whether these specific features were responsible for the improved accuracy rates. This approach was guided by the literature on distinct processing mechanisms for both internal and external features (e.g., Ellis et al., 1979; Young et al., 1985). Experiment 3 revealed that external facial features produced more reliable findings across all the analyses, with **AEFS A** outperforming both **AEFS C** and **AEFS D** in constructing accurate composites.

While the findings contradicted the initial hypothesis, they again shed light on the distinct differences in composite quality when implementing the H-CI as part of the investigative interview. This theme of composite quality variation is persistent throughout this thesis. As emphasised by the previous experiments, the findings suggest the potential to modify the composite construction process when combining a practitioner-led version with the H-CI, potentially reducing the amount of time and effort needed to produce an accurate composite.

In an effort to further improve the percentage of correct composite naming, especially in the light of the unexpected results from the H-CI, a new direction was proposed. This involved reducing the retention interval between encoding the target face and engaging in the investigative interview prior to the composite construction session. The rationale behind this was that a shorter retention interval might enhance participants retrieval of information about the target face by preventing factors such as memory decay (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Krouse, 1981; Murre & Dros, 2015; Shepherd et al., 1991; see Frowd et al., 2015 for a meta-analysis). This formulated the foundation for the design of Experiment 4.

Experiment 4

In Experiment 4, the primary objective closely aligned with the goal outlined in the preceding three experiments. The main aim of Experiment 4 was to replicate the methodology of Experiment 3, with a specific emphasis on the H-CI. An intriguing observation had emerged during Experiment 3, which contradicted expectations based on existing literature (e.g., Frowd, Bruce & Hancock, 2008; Frowd, Bruce, Ness et al., 2007; Frowd, Bruce, Smith et al., 2008; Frowd et al., 2004; Frowd, McQuiston-Surrett et al., 2005; Frowd, Nelson et al., 2012). Surprisingly, the implementation of the H-CI, alone, led to a notable decrease in the percentage of correct composite naming scores. In light of this unexpected outcome, a proposal was put forward to enhance the effectiveness of the H-CI by combining it with a significantly shortened retention interval. Unlike the previous 24-hour retention interval, Experiment 4 introduced a considerably shorter retention interval of 3-4 hours. This adjustment was grounded in the understanding that reducing the time between encoding and retrieval of information could potentially facilitate witnesses in recalling more accurate details (e.g., Frowd, Bruce, Ness et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett et al., 2005; Krouse, 1981; Murre & Dros, 2015; Shepherd et al., 1991; see Frowd et al., 2015 for a meta-analysis). The rationale behind this modification was to minimise the gap between the encoding and retrieval of information, ultimately aiming to improve witness memory recall accuracy and, consequently, the accuracy of the constructed composites (e.g., Murre & Dros, 2015).

Consistent with the patterns observed in the preceding experiments, there was no reliable difference in accuracy between the **AEFS** and the percentage of correctly named composites. Nevertheless, it became apparent that the percentage of correctly named composites increased to 26.8%, returning to a level closely resembling the accuracy achieved during the initial implementation of the CI technique. Although the increase in the percentage of correct composite naming scores was not substantial, it indicated a positive trend toward improvement.

The evaluation of the likeness task involved three distinct analyses, mirroring the methodology employed in Experiment 3. First, in the analysis of the whole face it was observed that **AEFS D** produced more accurate composites compared to those constructed using **AEFS B**. Notably, this finding demonstrated the impact of introducing the H-CI, along with a shortened retention interval, as the new Holistic Scales outperformed the original Holistic Scales. This observation supported the experimental hypothesis, suggesting that providing participants with increased opportunities to make RBE to the internal facial features of a composite leads to higher accuracy in their construction.

Regarding the analysis of the internal facial features, the results demonstrated that **AEFS D** produced more accurate composites than those constructed in **AEFS A**. Furthermore, both **AEFS C** and **AEFS D** outperformed, for accuracy, composites constructed in **AEFS B**. Once again, these findings aligned with the experimental hypothesis, indicating that the more opportunity that participants had to make RBE to the internal features of a facial composite helped increase the accuracy of the resulting composites.

When comparing these findings across the entirety of this thesis, Experiment 4 emerged as the first instance where the experimental hypothesis was met. The consistent pattern of results highlighted the inconsistency in composite quality, across different **AEFS**, under different experimental conditions. The incorporation of a shorter retention interval, combined with the H-CI, allowed the software to be used to its full potential. Notably, for the first time throughout the thesis, progressing comprehensively through the software produced the most optimal composite accuracy outcomes. These findings are beneficial for forensic practice, as it highlights the optimal combination of memory retrieval techniques, to produce the most accurate composites using the full procedure of EvoFIT.

Limitations and future research

Face to face research: The primary recommendation centres around improving the methodology employed in the experiments. A key limitation to note is that the experiments were conducted under constraints imposed by COVID-19 regulations, which required the refrainment of traditional face-to-face research methods, in a controlled laboratory. To address this limitation, it is suggested that the experiments be replicated within a controlled laboratory environment. This approach would provide researchers with greater control over variables, such as the environment (e.g., Elliot et al., 2022) allowing for better monitoring of factors such as distraction and motivation.

Another highlighted aspect of this recommendation is the potential advantages of conducting interactions in person, especially for assessing non-verbal cues, which are an essential component of establishing rapport, and are typically more perceptible in face-to-face interactions compared to video calls (Jiang, 2020). The aim of these adjustments is to establish a higher degree of control over the experimental environment (e.g., Elliot et al., 2022), potentially mitigating concerns such as anxiety, stress, distraction, and cognitive load, resulting in heightened accuracy in the composites.

Multiple testing over multiple retrieval intervals: Another potential direction would be to test different opportunities at retrieval, in concurrence with the multiple opportunities to make RBE (i.e., the current aim of this thesis). Participants could be split into groups, for example Group 1 of participants could be interviewed and construct the composite 24 hours after exposure to the target photograph, which is standard police procedure (Fodarella et al., 2015). A second group of participants could be interviewed at different, practical, retention intervals (e.g., 4 hours after exposure to the target photograph, and 24 hours after exposure, before constructing the composite). Finally, a third group could be interviewed at even more, again practical, time intervals, for example 4, 8 and 24 hours (before constructing the composite) after exposure to the target photograph. The three groups can then be compared to see if multiple retrieval attempts help to increase the overall quality of composites constructed. Specifically, if multiple retrieval attempts and the addition of multiple opportunities to make more RBE (i.e., **AEFS A**, **B**, **C**, and **D**) can increase overall composite quality. This could potentially be taken even further and could look at the participants self-administering the interview, as we know the police resources are limited (Fodarella et al., 2015) this could help to overcome the issues of lengthy retention intervals, whilst strengthening witnesses' memory. This procedure would involve the participants, at allotted time intervals, writing down their recall of the facial features. This could then be compared to the practitioner-led interviews to see if they have similar effects, or if one outweighs the other. If self-administered is better than practitioner-led, it could potentially be implemented to help increase quantity and quality of information that participants recall and increase composite quality without using police resources.

Documenting changes made: Another future recommendation that could be implemented regards documenting the number of changes that participants make, specifically at **AEFS C** and **AEFS D**. By doing this it would allow the researcher to identify whether the number of changes corresponds with the naming and likeness ratings, (i.e., do more changes increase the likelihood that composites will be named or does it increase the likeness ratings for similarity between the composite and the target photograph).

Compare retention intervals: An additional direction regards making a formal comparison between both Experiment 3 and Experiment 4. Experiment 3 focused on the H-CI with a longer retention interval of 24 hours, which is typically seen in practise (Fodarella et al., 2015). Where Experiment 4 focused on the H-CI with a shorter retention interval of 3-4 hours. For both experiments different stimuli was used (EastEnders and Coronation Street correspondingly) meaning a formal comparison could not be made. With this is mind it would be beneficial to compare both retention intervals (3-4 hours vs 24 hours) with the same stimuli to see if the retention interval impacted composite quality.

Task difficulty and anxiety: As task difficulty has been addressed throughout the thesis it may be worth considering asking participants to fill out a questionnaire after the task to document their own personal interpretation of how difficult they found the task of constructing a composite. This could be used in regard to compare the self-administered version of EvoFIT to the practitioner-led version to make a formal comparison between the two procedures and identify whether doing the task alone or with the help of a forensic practitioner influences task difficulty. There is also the potential for another questionnaire to be implemented to compare the self-administered version of EvoFIT versus the practitioner led. Alongside the questionnaire for task difficulty could be a questionnaire measuring participants level of anxiety. As there has not really been a formal comparison between the self-administered version and practitioner-led, it would be interesting to see whether this aligns with existing research (e.g., Martin et al., 2017; Martin et al., 2018; Risan et al., 2016) that constructing a composite from the comfort of your own home is less anxiety provoking than constructing a composite with a forensic practitioner in an unknown setting.

Integrating AEFS D: Finally, a noteworthy and positive suggestion, is to consider integrating the new **AEFS D** Stage into the composite software EvoFIT. Throughout the EvoFIT software, which was manipulated for this thesis, **AEFS D** functions as an optimal Stage where witnesses have the opportunity to make more RBE to the internal facial features of the composite, if they are unsatisfied with the results achieved through **AEFS C**, the usual Stage for saving and distributing composites. Importantly, **AEFS D** consistently produced more accurate composites, in regard to likeness, throughout the entirety of the thesis, compared to those constructed using **AEFS C**. Therefore, if the composite software is to be employed in practical contexts, incorporating **AEFS D** as an optional enhancement for witnesses seeking greater composite accuracy is a favourable suggestion. To test these findings further, it would be interesting to change the procedure from the typical **A-B-C-D** to other sequences (i.e. **AEFS A** straight to **AEFS D** or **A-B-D-C**) to see whether **AEFS D** alone increases composite accuracy, or whether **AEFS B** and **AEFS C** reduce composite quality and **AEFS D** just brings the quality back up.

General Conclusion

The core objective of this thesis was grounded in the theoretical underpinnings of enhancing eyewitness memory. The research demonstrated that strategically employing multiple retrieval attempts during investigative interviews can help witnesses to retrieve more accurate information (e.g., Brown et al., 2017; Butler & Roediger, 2007; Fisher & Geiselman, 1992; Geiselman et al., 1984; Roediger & Karpicke, 2006; see Köhnken et al., 1999; Memon, Meissner et al., 2010 for a meta-analysis). Importantly, this approach had not previously been applied to the procedure of constructing a composite, and it was proposed that integrating this technique would lead to more precise composites based on sights from the literature on multiple retrieval opportunities.

Furthermore, the thesis sought to investigate the potential advantages of using EvoFIT software, particularly in conjunction with memory facilitating techniques. The results indicated that composites constructed using the WAH software achieved the highest accuracy when constructed exclusively through the application of **AEFS A**, which covers the initial Stage of EvoFIT: evolving the face of the composite. Composites constructed using the CI produced most accurate composites through the application of **AEFS B**, involving the use of the original Holistic Scales of the software. For composites constructed using the H-CI approach, optimal accuracy was attained when composites were constructed through the application of **AEFS A**, reaffirming the importance of the initial Stages of EvoFIT.

It is important to note that the hypotheses predicting that providing witnesses with more opportunity to make RBE to the internal facial features of a composite would enhance its accuracy were contradicted by the findings of the initial three experiments. However, these results did reveal distinct differences within the EvoFIT software concerning composite quality and memory enhancing techniques (e.g., the CI and the H-CI). These observations highlight the potential for adaptions in the composite construction process based on the specific procedure, and memory facilitating techniques used. In contrast, Experiment 4 supported the experimental hypothesis, showing that a short retention interval, in combination with the H-CI, produced the most accurate composites, particularly when multiple opportunities to make RBE were integrated.

The findings throughout this thesis also shed light on the fact that the conventional point, at which composites are typically saved and distributed (**AEFS C**), does not consistently produce the most accurate composites under the WAH, CI, H-CI, and H-CI with a short retention interval, conditions. This insight sheds light on the potential reasons for the recurring issues found in witness misidentifications (e.g., Innocence Project, 2023). This discovery implies that a simple adaption to the software, specifically the use of **AEFS D**, could significantly enhance composite accuracy and potentially reduce errors in the criminal justice system (e.g., Innocence Project, 2023).

In summary, this thesis highlights the complex relationship between memory facilitating techniques, composite construction, and accuracy. The evaluation of these factors contributes to both the existing body of literature surrounding eyewitness memory and offers insight into the ongoing challenge of constructing accurate composites in practice.

REFERENCES

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, *22*(3), 261-295.

Anderson, K. J. (1994). Impulsivity, caffeine, and task difficulty: A within-subjects test of the Yerkes-Dodson law. *Personality and Individual Differences*, *16*(6), 813-829.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063-1087.

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of verbal learning and verbal behavior*, *17*(1), 1-12.

Baddeley, A. D. (1979). Working memory and reading. *Processing of visible language*, (pp. 355-370).

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Psychology Press.

Baddeley, A. D. (1998). Working memory. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie*, *321*(2-3), 167-173.

Baddeley, A. D. (2002). Is working memory still working?. *European psychologist*, *7*(2), 85-97.

Baddeley, A. D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval?. *Memory & Cognition*, *21*(2), 146-155.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*, *8*, 47–89.

Bell, M. C., Kawadri, N., Simone, P. M., & Wiseheart, M. (2014). Long-term memory, sleep, and the spacing effect. *Memory*, *22*(3), 276-283.

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.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *From learning processes to cognitive processes: Essays in honor of William K. Estes*, *2*, 35-67.

Bolton, S., & Robinson, O. J. (2017). The impact of threat of shock-induced anxiety on memory encoding and retrieval. *Learning & Memory*, *24*(10), 532-542.

Bornstein, B. H., Liebel, L. M., & Scarberry, N. C. (1998). Repeated testing in eyewitness memory: A means to improve recall of a negative emotional event. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *12*(2), 119-131.

Brace, N. A., Pike, G. E., Allen, P., & Kemp, R. I. (2006). Identifying composites of famous faces: Investigating memory, language and system issues. *Psychology, Crime & Law*, *12*(4), 351-366.

Brown, C., Frowd, C. D., & Portch, E. (2017). 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*(2), 248-265.

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

Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*(4-5), 514-527.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human perception and performance*, *26*(2), 527.

Carter, C. A., Bottoms, B. L., & Levine, M. (1996). Linguistic and socioemotional influences on the accuracy of children's reports. *Law and Human Behavior*, *20*(3), 335-358.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin*, *132*(3), 1-35.

Chae, Y. (2010). Application of laboratory research on eyewitness testimony. *Forensic Psychology Practice. 10*(3), 252-261.

Chan, J. C., & LaPaglia, J. A. (2011). The dark side of testing memory: repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: Applied*, *17*(4), 418-433.

Chan, J. C., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological science*, *20*(1), 66-73.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction*, *8*(4), 293-332.

Christianson, S. Å. (1992). Emotional stress and eyewitness memory: a critical review. *Psychological bulletin*, *112*(2), 284-309.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, *3*, 149-210.

Clifford, B. R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology*, *63*(3), 352-359.

Collins, R., Lincoln, R., & Frank, M. G. (2002). The effect of rapport in forensic interviewing. *Psychiatry, psychology and law*, *9*(1), 69-78.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology*, *72*(4), 629-637.

Davies, G., Van der Willik, P., & Morrison, L. J. (2000). Facial composite production: A comparison of mechanical and computer-driven systems. *Journal of Applied Psychology*, *85*(1), 119-124.

Deffenbacher, K. A. (2008). Estimating the impact of estimator variables on eyewitness identification: A fruitful marriage of practical problem solving and psychological theorizing. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *22*(6), 815-826.

Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once-seen face: estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied*, *14*(2), 139-150.

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

Delleman, B., & Fernandes, M. (2015). Individual differences in anxiety influence verbal memory accuracy and confidence. *Journal of Individual Differences*, *36*(2), 73-79.

Dobson, M., & Markham, R. (1992). Individual differences in anxiety level and eyewitness memory. *The Journal of general psychology*, *119*(4), 343-350.

Duggan, A. P., & Parrott, R. L. (2001). Physicians' nonverbal rapport building and patients' talk about the subjective component of illness. *Human Communication Research*, *27*(2), 299-311.

Duke, M. C., Wood, J. M., Bollin, B., Scullin, M., & LaBianca, J. (2018). Development of the Rapport Scales for Investigative Interviews and Interrogations (RS3i), Interviewee Version. *Psychology, Public Policy, and Law*, *24*(1), 64.

Eisenberg, P. (2019). The cognitive interview and enhanced cognitive interview in financial forensics and investigations. *Journal of Contemporary Research in Social Sciences*, *1*(1), 55-64.

Elliott, E. M., Bell, R., Gorin, S., Robinson, N., & Marsh, J. E. (2022). Auditory distraction can be studied online! A direct comparison between in-person and online experimentation. *Journal of Cognitive Psychology*, *34*(3), 307-324.

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.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of learning and motivation*, *44*, 145-200.

Eysenck, M. W. (1979). Anxiety, learning, and memory: A reconceptualization. *Journal of research in personality*, *13*(4), 363-385.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & emotion*, *6*(6), 409-434.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, *7*(2), 336-353.

Finger, K., & Pezdek, K. (1999). The effect of cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal of Applied Psychology, 84*(3), 340-348.

Fisher, R. P., & Geiselman, R. E. (1992). Memory enhancing techniques for investigative interviewing: The cognitive interview, (pp. 1-232)

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. (1987). Enhancing enhanced eyewitness memory: Refining the cognitive interview. *Journal of Police Science & Administration 15*(4), 291-297.

Fisher, R. P., Ross, S. J., & Cahill, B. S. (2017). Interviewing witnesses and victims. *Forensic psychology in context*, (pp. 56-74).

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. (2011). Eyewitnesses and the use and application of cognitive theory. In G. Davey (Ed.). *Introduction to Applied Psychology,* (pp. 267-289).

Frowd, C. D. (2021). Forensic facial composites. In *Methods, measures, and theories in eyewitness identification tasks*, (pp. 34-64).

Frowd, C. D., & Fields, S. (2011). Verbalization effects in facial composite production. *Psychology, Crime & Law*, *17*(8), 731-744.

Frowd, C. D., & Hancock, P. J. (2008). Evolving human faces. *The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music*, (pp. 189-210).

Frowd, C. D., Bruce, V., & Hancock, P. J. (2008). Changing the face of criminal identification. *The Psychologist*, *21*(8), 668-672.

Frowd, C. D., Bruce, V., McIntyre, A., & Hancock, P. (2007). The relative importance of external and internal features of facial composites. *British journal of psychology*, *98*(1), 61-77.

Frowd, C. D., Bruce, V., Ness, H., Bowie, L., Paterson, J., Thomson-Bogner, C., & Hancock, P. J. (2007). Parallel approaches to composite production: interfaces that behave contrary to expectation. *Ergonomics*, *50*(4), 562-585.

Frowd, C. D., Bruce, V., Ross, D., McIntyre, A., & Hancock, P. J. (2007). An application of caricature: how to improve the recognition of facial composites. *Visual cognition*, *15*(8), 954-984.

Frowd, C. D., Bruce, V., Smith, A. J., & Hancock, P. J. (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-Surrett, D., Richardson, J., Baldwin, H., & Hancock, P. (2005). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology*, *10*(1), 63-81.

Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., Mclanaghan, S., & Hancock, P. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law*, *11*(1), 33-52.

Frowd, C. D., Erickson, W. B., Lampinen, J. M., Skelton, F. C., McIntyre, A. H., & Hancock, P. J. (2015). A decade of evolving composites: regression-and meta-analysis. *Journal of Forensic Practice*, *17*(4), 319-334.

Frowd, C. D., Hancock, P. J., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on applied perception (TAP)*, *1*(1), 19-39.

Frowd, C. D., Hancock, P. J., Bruce, V., McIntyre, A. H., Pitchford, M., Atkins, R., & Sendrea, G. (2010). Giving crime the'evo': catching criminals using EvoFIT facial composites. In *2010 International Conference on Emerging Security Technologies*, (pp. 36-43). IEEE.

Frowd, C. D., Hancock, P. J., Bruce, V., Skelton, F. C., Atherton, C. J., Nelson, L., & Sendrea, G. (2011). Catching more offenders with EvoFIT facial composites: lab research and police field trials. *Global Journal of Human-Social Science, 11*(3), 35-46.

Frowd, C. D., McQuiston-Surrett, D., Anandaciva, S., Ireland, C. G., & Hancock, P. J. (2007). An evaluation of US systems for facial composite production. *Ergonomics*, *50*(12), 1987-1998.

Frowd, C. D., McQuiston-Surrett, D., Kirkland, I., & Hancock, P. (2005). The process of facial composite production. *Forensic Psychology and Law,* (pp. 140-152).

Frowd, C. D., Nelson, L., Skelton, F., Noyce, R., Atkins, R., Heard, P., …& Hancock, P. J. (2012). Interviewing techniques for Darwinian facial-composite systems. *Applied Cognitive Psychology, 26*(4), 576-584.

Frowd, C. D., Pitchford, M., Skelton, F., Petkovic, A., Prosser, C., & Coates, B. (2012). Catching even more offenders with EvoFIT facial composites. In *2012 Third International Conference on Emerging Security Technologies,* (pp. 20-26). IEEE.

Frowd, C. D., Portch, E., Killeen, A., Mullen, L., Martin, A. J., & Hancock, P. J. (2019). 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., Atherton, C., Pitchford, M., Hepton, G., Holden, L., & Hancock, P. J. (2012). Recovering faces from memory: The distracting influence of external facial features. *Journal of Experimental Psychology: Applied, 18*(2), 224-265.

Frowd, C. D., Skelton, F., Hepton, G., Holden, L., Minahil, S., Pitchford, M., & Hancock, P. J. (2013). Whole-face procedures for recovering facial images from memory. *Science & Justice*, *53*(2), 89-97.

Geiselman, R. E., Fisher, R. P., Firstenberg, I., Hutton, L. A., Sullivan, S. J., Avettissian, I. and Prosk, A. (1984). Enhancement of eyewitness memory: an empirical evaluation of the cognitive interview. *Journal of Police Science and Administration, 12*(1), 74-80.

Geiselman, R.E., Fisher, R.P., MacKinnon, D.P. and Holland, H.L. (1985), “Eyewitness memory enhancement in the police interview: cognitive retrieval mnemonics versus hypnosis”, *Journal of Applied Psychology, 70*(2), 401-12.

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.

Griffiths, A., & Milne, B. (2010). The application of cognitive interview techniques as part of an investigation. *Consultancy and advising in forensic practice: Empirical and practical guidelines*, (pp. 69-90).

Guzel, M. A., & Higham, P. A. (2013). Dissociating early-and late-selection processes in recall: The mixed blessing of categorized study lists. *Memory & Cognition*, *41*, 683-697.

Hanway, P., Akehurst, L., Vernham, Z., & Hope, L. (2021). The effects of cognitive load during an investigative interviewing task on mock interviewers’ recall of information. *Legal and Criminological Psychology*, *26*(1), 25-41.

Henkel, L. A. (2004). Erroneous memories arising from repeated attempts to remember. *Journal of Memory and Language*, *50*(1), 26-46.

Het, S., Ramlow, G., & Wolf, O. T. (2005). A meta-analytic review of the effects of acute cortisol administration on human memory. *Psychoneuroendocrinology*, *30*(8), 771-784.

Hine, K., & Itoh, Y. (2018). Reducing the negative effect of the retention interval on the composite face recognition. *The Journal of General Psychology*, *145*(3), 296-312.

Hintzman, D. L., & Rogers, M. K. (1973). Spacing effects in picture memory. *Memory & Cognition*, *1*(4), 430-434.

Innocence Project. (2023). *Explore the Numbers: Innocence Project’s Impact.* Retrieved from: https://innocenceproject.org/exonerations-data/.

Iriye, H., & Jacques, P. L. S. (2020). How visual perspective influences the spatiotemporal dynamics of autobiographical memory retrieval. *Cortex*, *129*, 464-475.

Jiang, M. (2020). The reason Zoom calls drain your energy. *BBC Worklife*, *23*.

Kieckhaefer, J. M., Vallano, J. P., & Schreiber Compo, N. (2014). Examining the positive effects of rapport building: When and why does rapport building benefit adult eyewitness memory?. *Memory*, *22*(8), 1010-1023.

Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime and Law*, *5*(1-2), 3-27.

Krouse, F. L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology*, *66*(5), 651-654.

Laughery, K R., Fessler, P. K., Lenorovitz, D. R., & Yoblick, D. A. (1974). Time delay and similarity effects in facial recognition. *Journal of Applied Psychology*, *59*(4), 490-496.

Laughery, K. R., Duval, C., & Wogalter, M. S. (1986). Dynamics of facial recall. In *Aspects of face processing*, (pp. 373-387).

Lee, H. C., & Pagliaro, E. M. (2013). Forensic evidence and crime scene investigation. *Journal of Forensic Investigation*, *1*(2), 1-5.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

Loftus, E. (2003). Our changeable memories: Legal and practical implications. *Nature Reviews Neuroscience*, *4*(3), 231-234.

MacLeod, M. (2002). Retrieval‐induced forgetting in eyewitness memory: Forgetting as a consequence of remembering. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *16*(2), 135-149.

Marr, C., Sauerland, M., Otgaar, H., Quaedflieg, C. W., & Hope, L. (2021). The effects of acute stress on eyewitness memory: An integrative review for eyewitness researchers. *Memory*, *29*(8), 1091-1100.

Martin, A. J., Hancock, P. J., & Frowd, C. D. (2017). 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.

Martin, A. J., Hancock, P. J., Frowd, C. D., Heard, P., Gaskin, E., Ford, C., & Hewett, T. (2018). EvoFIT composite face construction via practitioner interviewing and a witness-administered protocol. In *2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, (pp. 311-316). IEEE.

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: an example of the two-way street between cognition and instruction. *New directions for teaching and learning*, *2002*(89), 55-71.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, *38*(1), 43-52.

Mcdermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition*, *34*(2), 261-267.

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(6), 603-616.

Memon, A., & Gawrylowicz, J. (2006). The cognitive interview. *The handbook of communication skills*, (pp. 541-560).

Memon, A., & Higham, P. A. (1999). A review of the cognitive interview. *Psychology, Crime and Law*, *5*(1-2), 177-196.

Memon, A., & Koehnken, G. (1992). Helping witnesses to remember more: The cognitive interview. *Expert Evidence*, *1*(2), 39-48.

Memon, A., Bull, R., & Smith, M. (1995). Improving the quality of the police interview: Can training in the use of cognitive techniques help?. *Policing and Society: An International Journal*, *5*(1), 53-68.

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, public policy, and law*, *16*(4), 340-372.

Memon, A., Wark, L., Bull, R., & Koehnken, G. (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology*, *88*(2), 179-197.

Memon, A., Zaragoza, M., Clifford, B. R., & Kidd, L. (2010). Inoculation or antidote? The effects of cognitive interview timing on false memory for forcibly fabricated events. *Law and Human Behavior*, *34*(2), 105-117.

Milne, R. J. (1997). *Analysis and application of the cognitive interview* (Doctoral dissertation, University of Portsmouth).

Milne, R., Bull, R., Koehnken, G., & Memon, A. (1995). The Cognitive Interview and suggestibility. *Issues in Criminological & Legal Psychology*.

Mostyn, G. R. (2012). Cognitive load theory: What it is, why it's important for accounting instruction and research. *Issues in Accounting Education*, *27*(1), 227-245.

Mueller, J. H. (1979). Anxiety and encoding processes in memory. *Personality and Social Psychology Bulletin*, *5*(3), 288-294.

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: a meta-analytic review of retrieval-induced forgetting. *Psychological bulletin*, *140*(5), 1383-1409.

Murre, J. M., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLoS ONE*, 10(7), e0120644.

Nahouli, Z., Dando, C. J., Mackenzie, J. M., & Aresti, A. (2021). Rapport building and witness memory: Actions may 'speak'louder than words. *PLoS ONE*, *16*(8), e0256084.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5-6), 389-395.

Nash, R. A., Nash, A., Morris, A., & Smith, S. L. (2016). Does rapport-building boost the eyewitness eyeclosure effect in closed questioning?. *Legal and Criminological Psychology*, *21*(2), 305-318.

Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy–confidence relation in eyewitness memory. Applied Cognitive Psychology, 20(7), 973-985.

Odinot, G., Memon, A., La Rooy, D., & Millen, A. (2013). Are two interviews better than one? Eyewitness memory across repeated cognitive interviews. *PloS one*, *8*(10), e76305.

Odinot, G., Wolters, G., & Lavender, T. (2009). Repeated partial eyewitness questioning causes confidence inflation but not retrieval-induced forgetting. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(1), 90-97.

Office for National Statistics. (2023). *Crime in England and Wales: year ending March 2023.* Retrievedfrom: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingmarch2023

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, *38*(1), 1-4.

Paivio, A. (1986). Mental representations: A dual-coding approach. *New York: Oxford University Press,* (pp. 1-322).

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255-287.

Paivio, A., & Clark, J. M. (2006). Dual coding theory and education. *Pathways to literacy achievement for high poverty children*, (pp. 1-20).

Panneerchelvam, S., & Norazmi, M. N. (2003). Forensic DNA profiling and database. *The Malaysian journal of medical sciences: MJMS*, *10*(2), 20-26.

Paulo, R. M., Albuquerque, P. B., & Bull, R. (2013). The enhanced cognitive interview: Towards a better use and understanding of this procedure. *International Journal of Police Science & Management*, *15*(3), 190-199.

Pitchford, M., Green, D., & Frowd, C. D. (2017). The impact of misleading information on the identifiability of feature-based facial composites. In *2017 Seventh International Conference on Emerging Security Technologies (EST)* (pp. 185-190). IEEE.

Raaijmakers, J. G., & Shiffrin, R. M. (1992). Models for recall and recognition. *Annual review of psychology*, *43*(1), 205-234.

Richardson, B. H., & Nash, R. A. (2022). 'Rapport myopia'in investigative interviews: Evidence from linguistic and subjective indicators of rapport. *Legal and Criminological Psychology*, *27*(1), 32-47.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological science*, *22*(4), 464-471.

Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces happens at a glance. *Vision research*, *49*(23), 2856-2861.

Risan, P., Binder, P. E., & Milne, R. (2016). Regulating and coping with distress during police interviews of traumatized victims. *Psychological trauma: theory, research, practice, and policy*, *8*(6), 736-744.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences*, *15*(1), 20-27.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological science, 17(3), 249-255.

Roediger, H. L., & Marsh, E. J. (2009). False memory. *Scholarpedia*, *4*(8), 3858.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of experimental psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Roediger, H. L., Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language*, *35*(2), 300-318.

Roediger, H. L., McDermott, K. B., & Goff, L. M. (1997). Recovery of true and false memories: Paradoxical effects of repeated testing. *Recovered Memories and False Memories*, (pp.118-149).

Roediger, H. L., Zaromb, F. M., & Butler, A. C. (2009). The role of repeated retrieval in shaping collective memory. *Memory in mind and culture*, (pp. 138-170).

Sauro, M. D., Jorgensen, R. S., & Teal Pedlow, C. (2003). Stress, glucocorticoids, and memory: a meta-analytic review. *Stress*, *6*(4), 235-245.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational psychology review*, *19*, 469-508.

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.

Schweitzer, K., & Nuñez, N. (2021). The effect of evidence order on jurors' verdicts: Primacy and recency effects with strongly and weakly probative evidence. *Applied Cognitive Psychology*, *35*(6), 1510-1522.

Scrivner, E., & Safer, M. A. (1988). Eyewitnesses show hypermnesia for details about a violent event. *Journal of Applied Psychology*, *73*(3), 371-377.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological bulletin*, *100*(2), 139-156.

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic bulletin & review*, *2*, 249-253.

Shepherd, J. W., & Ellis, H. D. (1973). The effect of attractiveness on recognition memory for faces. *The American journal of psychology*, 627-633.

Shepherd, J. W., & Ellis, H. D. (1996). Face recall—Methods and problems. In *Psychological issues in eyewitness identification* (pp. 87-115).

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.

Shields, G. S., Moons, W. G., Tewell, C. A., & Yonelinas, A. P. (2016). The effect of negative affect on cognition: Anxiety, not anger, impairs executive function. *Emotion*, *16*(6), 792-811.

Siegel, J. M., & Loftus, E. F. (1978). Impact of anxiety and life stress upon eyewitness testimony. *Bulletin of the Psychonomic Society*, *12*(6), 479-480.

Smith, A. M., Floerke, V. A., & Thomas, A. K. (2016). Retrieval practice protects memory against acute stress. *Science*, *354*(6315), 1046-1048.

St-Yves, M. (2006). The psychology of rapport: Five basic rules. *Investigative interviewing*, 82-106.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257-285.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances, (pp. 29-47).

Sweller, J., Ayres, P., Kalyuga, S., (2011). Cognitive load theory. In *Psychology of learning and motivation*, *55,* 37-76.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of experimental psychology: general*, *119*(2), 176-192.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225-245.

Thomas, A. K., Bulevich, J. B., & Chan, J. C. (2010). Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *Journal of Memory and Language*, *63*(2), 149-157.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological inquiry*, *1*(4), 285-293.

Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied*, *9*(2), 101-118.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*(5), 352.

Valentine, T., & Mesout, J. (2009). Eyewitness identification under stress in the London Dungeon. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(2), 151-161.

Vallano, J. P., & Schreiber Compo, N. S. (2011). A comfortable witness is a good witness: Rapport-building and susceptibility to misinformation in an investigative mock-crime interview. *Applied cognitive psychology*, *25*(6), 960-970.

Vallano, J. P., & Schreiber Compo, N. S. (2015). Rapport-building with cooperative witnesses and criminal suspects: A theoretical and empirical review. *Psychology, Public Policy, and Law*, *21*(1), 85-99.

Vredeveldt, A., Hitch, G. J., & Baddeley, A. D. (2011). Eyeclosure helps memory by reducing cognitive load and enhancing visualisation. *Memory & cognition*, *39*, 1253-1263.

Warrington, E. K., & Ackroyd, C. (1975). The effect of orienting tasks on recognition memory. *Memory & Cognition*, *3*, 140-142.

Wells, G. L., & Hasel, L. E. (2007). Facial composite production by eyewitnesses. *Current Directions in Psychological Science*, *16*(1), 6-10.

Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory & Cognition*, *12*(4), 338-344.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual review of Psychology*, *54*(1), 277-295.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological science in the public interest*, *7*(2), 45-75.

Wells, G. L., Olson, E. A., & Charman, S. D. (2002). The confidence of eyewitnesses in their identifications from lineups. *Current Directions in Psychological Science*, *11*(5), 151-154.

Wheeler, M. A., & Roediger III, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3*(4), 240-246.

Winograd, E. (1976). Recognition memory for faces following nine different judgments. *Bulletin of the Psychonomic Society*, *8*(6), 419-421.

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(3), 181-190.

Wixted, J. T. (2018). Time to exonerate eyewitness memory. *Forensic science international*, *292*, 13-15.

Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the reliability of eyewitness memory. *Perspectives on Psychological Science*, *13*(3), 324-335.

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.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception, 14*(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., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of applied psychology*, *71*(2), 291-301.

Zahradnikova, B., Duchovicova, S., & Schreiber, P. (2018). Facial composite systems. *Artificial Intelligence Review*, *49*, 131-152.

Zaragoza, M. S., Belli, R. F., & Payment, K. E. (2013). Misinformation effects and the suggestibility of eyewitness memory. In *Do justice and let the sky fall* (pp. 35-63).

APPENDICES

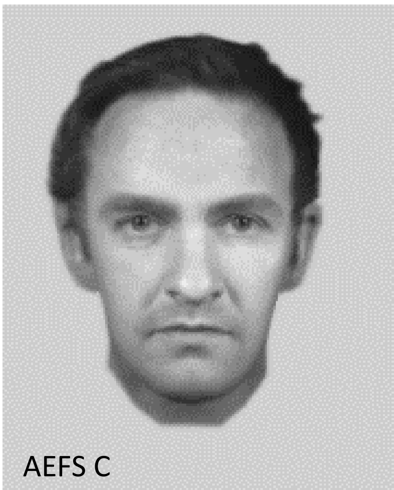
APPENDIX 1: EXPERIMENT 1 FACIAL COMPOSITES

Images have been resized for the purpose of the layout in the Thesis and do not accurately represent the format used throughout the experiments.

PETER BARLOW

AEFS A

AEFS B



AEFS C

AEFS D

TRACEY BARLOW

LEANNE BATTERSBY

CARLA CONNOR

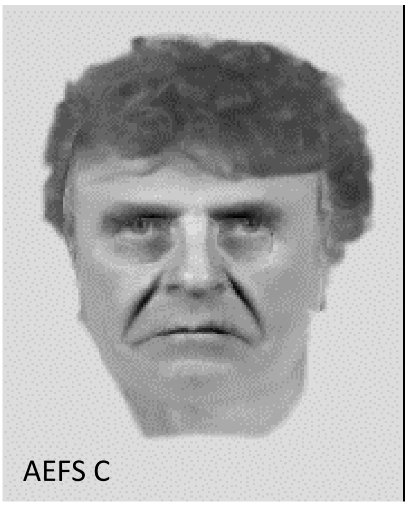
AEFS A

AEFS B

AEFS C

AEFS D

ROY CROPPER



TYRONE DOBBS

AEFS A

AEFS B

AEFS C

AEFS D

SALLY METCALFE

DAVID PLATT

KIRK SUTHERLAND

AEFS A

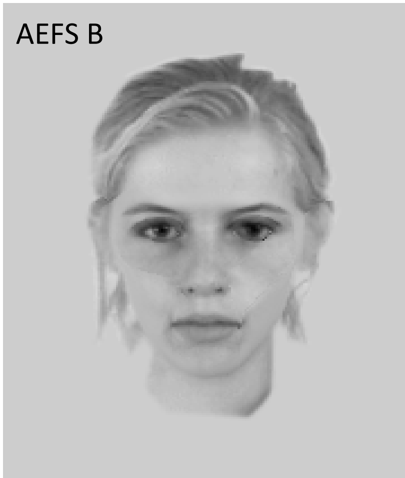
AEFS B

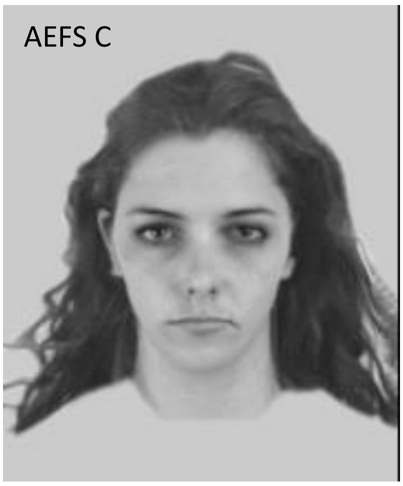
AEFS C

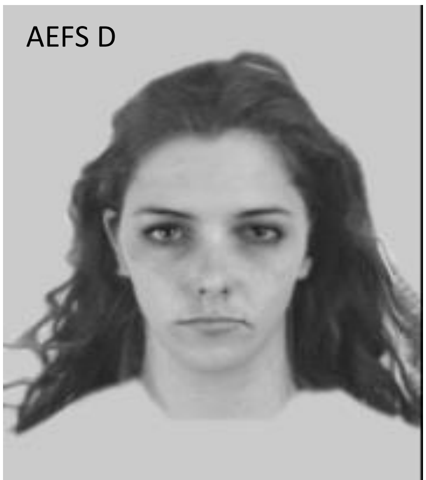
AEFS D

SOPHIE WEBSTER


AEFS A


AEFS B


AEFS C


AEFS D

APPENDIX 2: EXPERIMENT 2 FACIAL COMPOSITES

Images have been resized for the purpose of the layout in the Thesis and do not accurately represent the format used throughout the experiments.

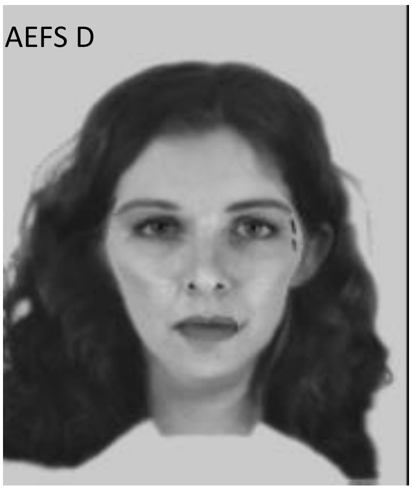
PETER BARLOW

TRACEY BARLOW

AEFS A

AEFS B

AEFS C



AEFS D

LEANNE BATTERSBY

CARLA CONNOR

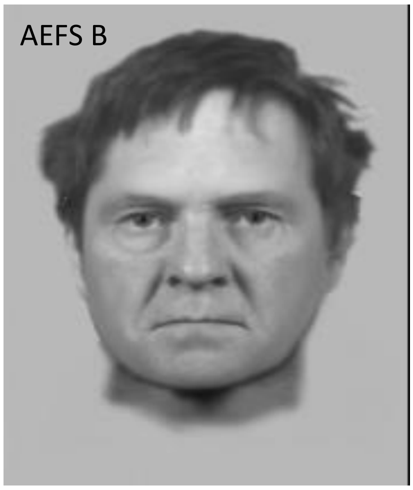

AEFS A


AEFS B


AEFS C


AEFS D

ROY CROPPER



TYRONE DOBBS

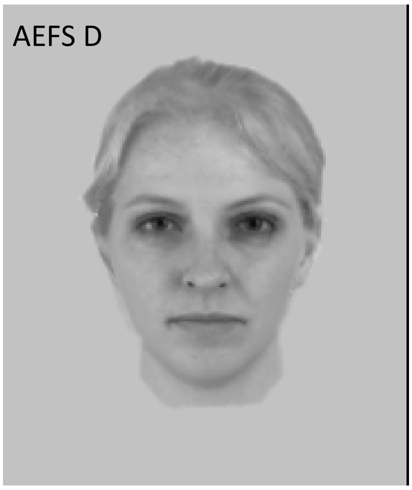
AEFS A

AEFS B

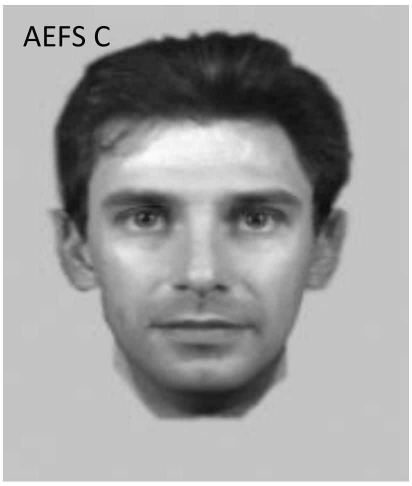
AEFS C

AEFS D

SALLY METCALFE

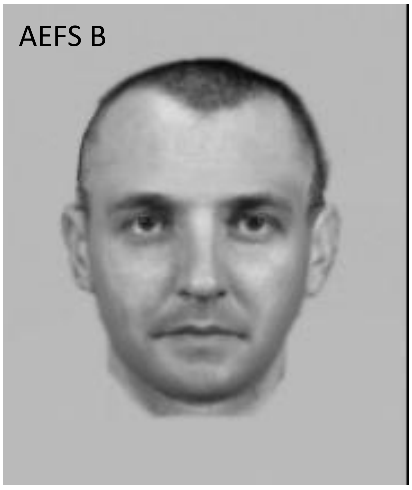


DAVID PLATT

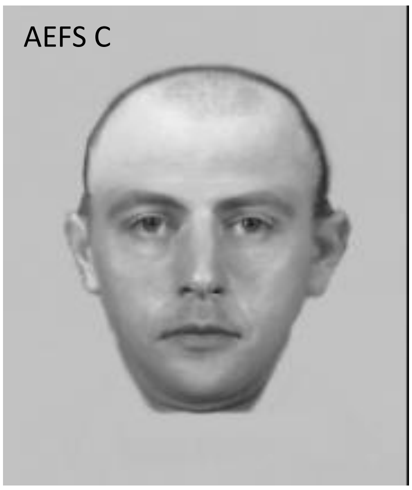


KIRK SUTHERLAND

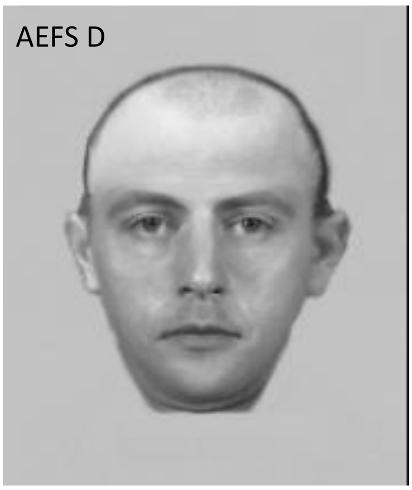
AEFS A



AEFS B



AEFS C



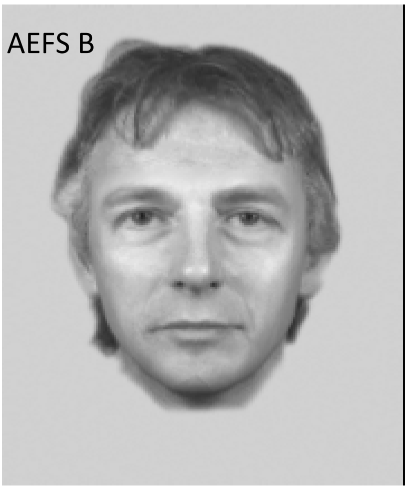
AEFS D

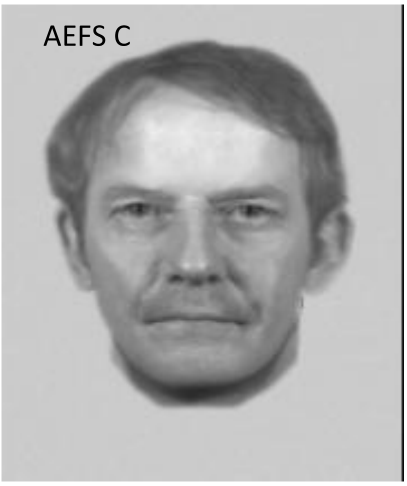
SOPHIE WEBSTER

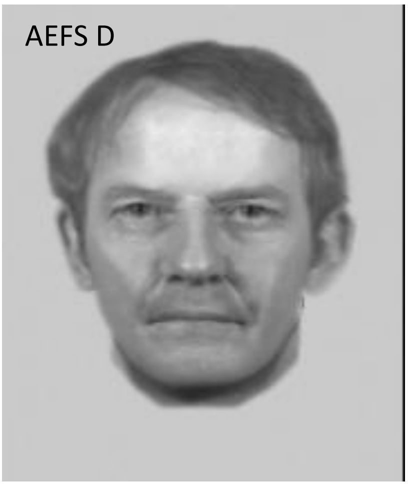
APPENDIX 3: EXPERIMENT 3 FACIAL COMPOSITES

Images have been resized for the purpose of the layout in the Thesis and do not accurately represent the format used throughout the experiments.

IAN BEALE







JANE BEALE

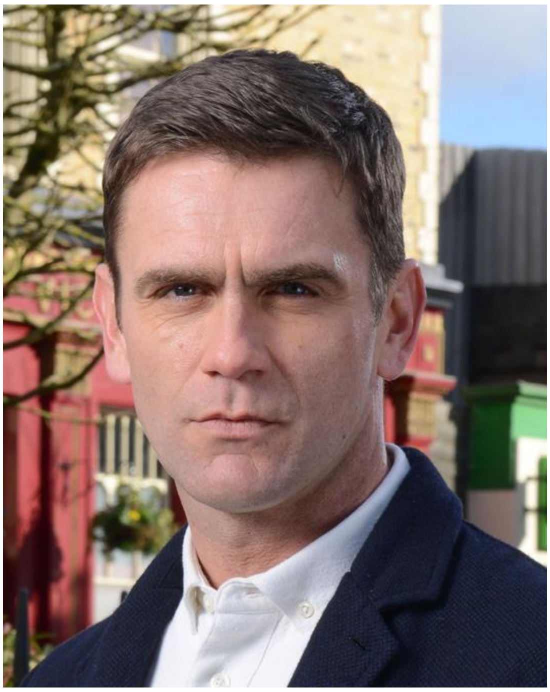
AEFS A

AEFS B

AEFS C

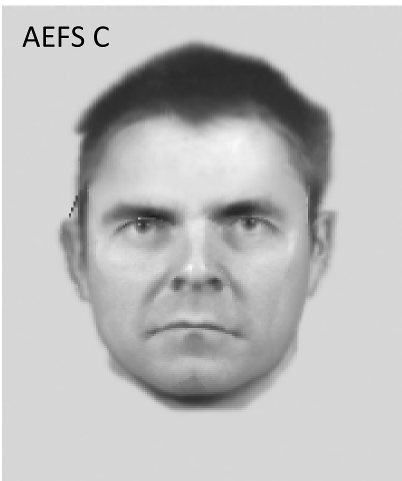
AEFS D

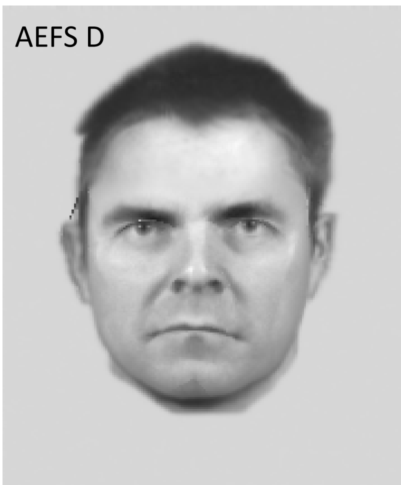
JACK BRANNING




AEFS A


AEFS B


AEFS C

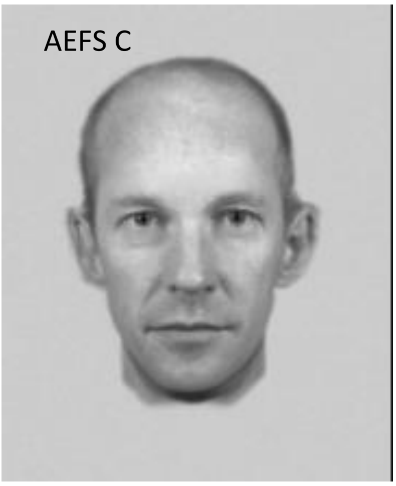

AEFS D

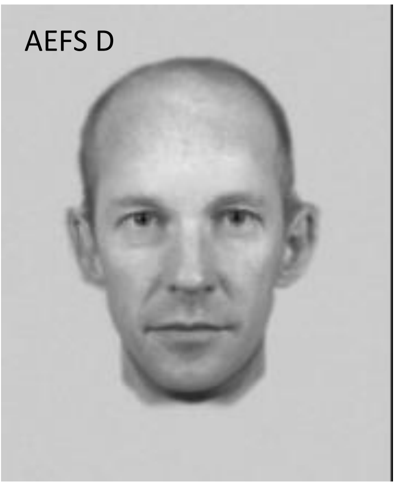
LAUREN BRANNING

MAX BRANNING


AEFS A


AEFS B


AEFS C


AEFS D

SHIRLEY CARTER

MARTIN FOWLER

AEFS A

AEFS B

AEFS C

AEFS D

BILLY MITCHELL

JEAN SLATER


AEFS A


AEFS B


AEFS C


AEFS D

STACEY SLATER

AEFS A

AEFS B

AEFS C

AEFS D

APPENDIX 4: EXPERIMENT 4 FACIAL COMPOSITES

Images have been resized for the purpose of the layout in the Thesis and do not accurately represent the format used throughout the experiments.

PETER BARLOW


AEFS A


AEFS B


AEFS C


AEFS D

TRACEY BARLOW


AEFS A


AEFS B


AEFS C


AEFS D

LEANNE BATTERSBY

AEFS A

AEFS B

AEFS C

AEFS D

FIZ BROWN

ROY CROPPER


AEFS A


AEFS B


AEFS C


AEFS D

EILEEN GRIMSHAW


AEFS A


AEFS B


AEFS C


AEFS D

STEVE MCDONALD


AEFS A


AEFS B


AEFS C


AEFS D

SALLY METCALFE

DAVID PLATT


AEFS A


AEFS B


AEFS C


AEFS D

KIRK SUTHERLAND

APPENDIX 5: VERBAL RECALL SHEET

Overall

Shape

Hair

Eyebrows

Eyes

Nose

Mouth

Ears