The Use of Facial Composites in Person Identification

Emma Garcia-Solley

School of Psychology

University of Kent

A thesis submitted for the degree of Ph.D. in the Faculty of Social Sciences at the University of Kent

Abstract

In criminal investigations, eyewitnesses are often required to construct a facial composite of a previously seen perpetrator to help the police narrow down a pool of suspects. Although research settings have previously indicated that composite identification rates are often low, the development of more recent holistic composite systems has substantially improved naming rates. However, considerably less research has explored other means of improving composite identification. This thesis investigated three strands of research for this purpose. Firstly, it was necessary to determine whether composite construction is a suitable task for eyewitnesses. In Chapter 2, the recognition accuracy of composite constructors and control participants was tested over forensically relevant delays, to determine whether any adverse effects of construction are likely to extend to applied settings. A null effect of composite construction on subsequent line-up performance was found across both a two-day, and ten-week time delay. Following this, ways of improving composite naming rates were investigated. In Chapter 3, the role of individual differences in composite construction was explored, to establish whether some eyewitnesses are better equipped for the task, based on face recognition abilities. In this instance, an individual's performance on two standardised recognition tests did not correspond to the quality of their facial composite. Finally, the role of context in composite identification was investigated. Chapter 4 suggests that composite naming can be increased substantially when contextual information is provided. These results are discussed within the context of applied settings and directions for future research are provided.

Acknowledgements

I would like to thank my supervisor, Dr. Markus Bindemann for agreeing to supervise me in unforeseen circumstances, and for providing excellent supervision throughout the duration of my PhD. I would also like to thank my husband, Jon, who supported me in my decision to return to education and provided much needed encouragement throughout.

Declaration

I declare that this thesis is my own work carried out under the normal terms of supervision.

Emma Garcia-Solley

Publications

Within this thesis, Chapter 2 has been presented at two conferences.

Chapter 2

Garcia, E., & Bindemann, M. (2017). Facial Composites: Helpful or Harmful? The Effect of Building a Facial Composite Over a Forensically Relevant Delay. Poster presented at the July meeting of the Experimental Psychology Society, University of Reading, UK.

Garcia, E., & Bindemann, M. (2017). Facial Composites: Helpful or Harmful? The Effect of Building a Facial Composite Over a Forensically Relevant Delay. Poster presented at the Innovations in Forensic Psychology Conference, University of Kent, UK.

Table of Contents

Chapter 1

General Introduction

1.1 Introduction

Criminal offences are a prevalent occurrence in the UK. Crime statistics for England and Wales show that 5.2 million offences were recorded in 2017, yielding a 13% increase in police recorded crime (Office for National Statistics, 2017). In many of these cases, eyewitnesses often play a pivotal role and are routinely required to identify a previously seen perpetrator in an identification procedure comprised of similar looking suspects (Police And Criminal Evidence Act, 1984). There are a number of identification techniques currently used by UK police forces (e.g. street ID, video ID, ID parade and group ID), which vary in frequency, but all fundamentally depend on the accuracy of an eyewitness (Davis, Valentine, Memon, & Roberts, 2015; Police And Criminal Evidence Act, 1984).

Despite the reliance of the justice system on eyewitness testimony, the process is error-prone (for reviews see, Wells, Memon, & Penrod, 2006; Wells & Olson, 2003). One insight into the scale of these errors is provided by data from the United States, where eyewitness *mis*identifications have contributed to 70% of wrongful convictions that have been overturned subsequently through DNA testing (Innocence Project, 2016). In the context of tens of thousands of identification line-ups being administered in the UK alone every year (Viper, 2009), the data of such DNA exoneration cases suggests that the inaccuracy of eyewitness testimony may present a large-scale, applied problem. This thesis focuses on one specific type of eyewitness testimony, comprising the use of facial composites.

1.2 Facial Composite Construction

When an eyewitness observes a crime but the identity and whereabouts of the perpetrator are unknown, the eyewitness may be asked by police to produce a criminal facial composite. These composites are visual representations of a perpetrator's face that are generated from an eyewitness' memory with specialist software (for reviews see, Davies &

Valentine, 2007; Frowd, 2017). Composites can be released to the public or shown to informants in the hope that potential suspects can be identified. To determine whether a suspect is, in fact, the perpetrator, these persons can then be presented back to the eyewitness in a police line-up. Focusing police inquiries through composite construction has proved to be indispensable in locating and apprehending perpetrators, with facial composites helping to solve hundreds of crimes, both in the UK and overseas (Frowd, Pitchford, et al., 2012; Solomon, Gibson, & Maylin, 2012).

Despite this large-scale usage, historically eyewitnesses have often found it difficult to construct facial composites that provide a good representation of a target face (Brace, Pike, Allen, & Kemp, 2006; Christie & Ellis, 1981; Frowd, Carson, Ness, Richardson, et al., 2005; Frowd, Bruce, & Hancock, 2008). This difficulty is reflected in past psychological studies of composite recognition, which typically generate low naming rates of less than 20% (Davies, van der Willik, & Morrison, 2000; Frowd, McQuiston-Surrett, Anandaciva, Ireland, & Hancock, 2007; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Valentine, Davis, Thorner, Solomon, & Gibson, 2010). The poor resemblance that has historically been depicted in facial composites is problematic, as eyewitness errors during this process can be detrimental to police investigations and, in extreme circumstances, contribute to wrongful convictions (Wells & Hasel, 2007).

In related forensic identification domains, such as passport control, unfamiliar face identification is much more difficult than familiar recognition (Hancock, Bruce, & Mike Burton, 2000; White, Kemp, Jenkins, Matheson, & Burton, 2014). By contrast, the difficulty experienced with composite construction extends to both familiar and unfamiliar faces, with a poor likeness of the target often being generated regardless of whether the suspect is known well to the witness, or whether only a brief encounter has taken place (Wells & Hasel, 2007). This is evident through low naming rates of composites created of unfamiliar famous faces

(Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005) and familiar faces known personally to constructors (Kovera, Penrod, Pappas, & Thill, 1997).

Figure 1.1. Celebrity composites made from memory of (from left to right, top to bottom) Brad Pitt, Graham Norton, Nicholas Cage, Michael Owen, Robbie Williams, Anthony (Ant) McPartlin, David Beckham and Noel Gallagher. Taken from Frowd et al (2008).

1.2.1 Generations of Composite Systems

The visual dissimilarity that can exist between target faces and composites has resulted in a variety of facial composite systems being invented, with each new system attempting to rectify previous limitations. From this, four generations of composite systems have evolved (for reviews see Davies & Valentine, 2007; Frowd, 2017). Traditionally, sketch artistry was used, whereby an artist drew a visual image of a suspect based on an eyewitness' verbal description (see bottom line of Figure 1.1). This traditional method was then replaced by second generation mechanical systems, which were used to construct a face through the build-up of individual facial features on acetates or cardboard, akin to the assembling of a puzzle (e.g. Identikit and Photofit; Davies & Valentine, 2007). Although mechanical systems allowed composites to be created by less artistic persons (Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005), composite quality was poor, even when composites were created with the target in view (Ellis, Davies, & Shepherd, 1978). Third generation software systems were subsequently developed (e.g. E-FIT and PROfit, see top two rows of Figure 1.1) and were guided by similar principles. As opposed to previous mechanical manipulations, however, computer software was used to manipulate images of the target face (Davies & Valentine, 2007). An initial improvement in composite quality was evident through the development of third generation systems, but quality decreased when composites were constructed from memory, suggesting such systems are unsuited to applied settings (Koehn & Fisher, 1997).

Finally, a fourth generation of composite systems was created (e.g. EvoFIT, see third line of Figure 1.1). These systems adopted an approach that relies on face recognition rather than face recall. Recall is thought to be a more difficult cognitive task than recognition (Davis, Sutherland, & Judd, 1961; Postman, Jenkins, & Postman, 1948), and eyewitnesses are often unable to recall a perpetrator's individual facial features in sufficient detail. Constructing a feature-based composite can therefore be extremely difficult, as the eyewitness is required to select individual facial features from an array of examples. By contrast, recognition is easier (Andrew & Bird, 1938; Postman et al., 1948).

Fourth generation systems provide an alternative approach by aligning composite construction more closely with face recognition (Frowd, 2017). Faces are thought to be encoded as whole entities, with facial features being more recognisable in the context of an entire face (Frowd et al., 2012; Tanaka & Farah, 1993; Wells & Hryciw, 1984). Fourth generation composite systems capitalise on this holistic process and require eyewitnesses to make selections from whole-face arrays. Consequently, these systems are often referred to as holistic systems.

There are currently three holistic systems in forensic use: EFIT-V (Gibson, Solomon, & Pallares-Bejarano, 2003), ID (Tredoux, Nunez, Oxtoby, & Prag, 2007) and EvoFIT (Frowd, Hancock, & Carson, 2004). Each of these systems present eyewitnesses with more than one face simultaneously and requires them to select the faces that most accurately resemble the perpetrator (see Figure 1.2). Using principle components analysis (PCA) and an evolutionary algorithm, the selections are then combined together to produce more identities, based on the previous choices (for reviews, see Frowd et al., 2004; Solomon et al., 2012). With EvoFIT, composite constructors are required to select six faces from a larger set of 18 whole-face arrays, across a number of screens and generations (see Figure 1.2). At the end of each generation, constructors are asked to select a single image that represents the best overall likeness to the target (see Figure 1.3). This image is subsequently given additional weighting in the evolutionary algorithm to improve target similarity (Frowd et al., 2004).

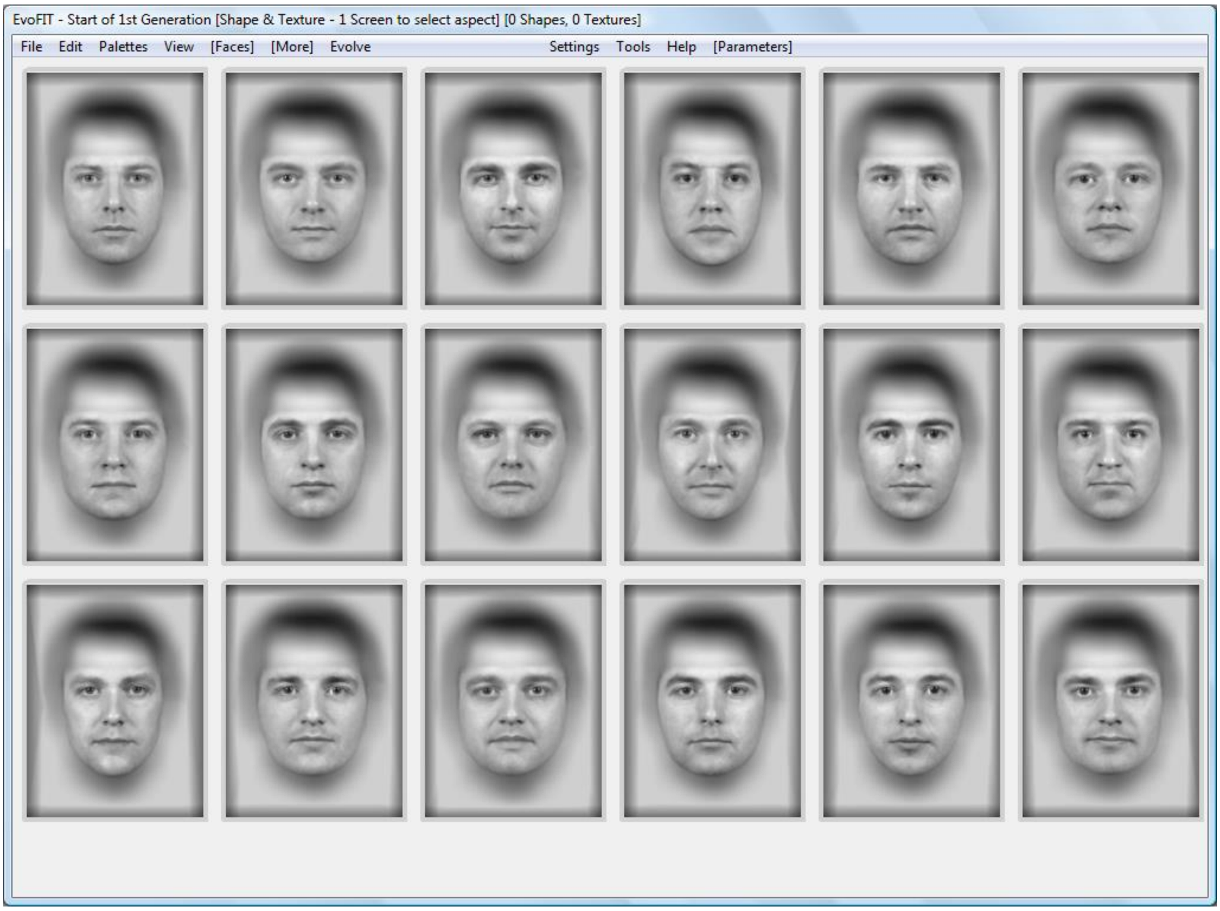

Figure 1.2. An example of an EvoFIT screen. Eyewitnesses are required to select two faces per screen across three generations. The external features are blurred to allow an eyewitness to focus on the internal features of the face. Taken from Frowd (2012).

Figure 1.3. An example of EvoFIT selections. At the end of each generation, eyewitnesses are required to select the best overall likeness to the target face. This sequence shows the faces selected at the end of each generation during composite construction of the singer, Robbie Williams. Taken from Frowd et al. (2009).

1.2.2 Naming Rates

The true value of a composite is determined by the rate it is correctly named. In applied settings, composites are created of unfamiliar faces in the hope that the visual representation will be recognised and named, allowing police to apprehend perpetrators of crime. The past decade of research has suggested that, when a short retention interval is employed between initial exposure to a target, and subsequent composite construction, a moderate likeness can be achieved. In such studies, composites created on feature-based systems elicit correct naming rates approximately 20% of the time (e.g. Bruce, Ness, Hancock, Newman, & Rarity, 2002; Davies et al., 2000). However, when a longer retention interval of two days is employed between viewing a target face and constructing a facial composite, naming rates of feature-based composites decrease even further to just 3% (Frowd et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005).

The decrease in naming rates over a prolonged interval between encoding and composite construction is a disconcerting finding. In applied settings, witnesses are contacted as soon as practical to allow their level of recall and suitability for composite construction to be assessed. Depending on a witness' availability, stress, and trauma levels, the interval between witnessing a crime and composite construction can vary, although two day delays are typical of criminal investigations (Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Frowd, Carson, Ness, Richardson, et al., 2005).

Despite the disappointingly low naming rates associated with feature-based systems, the evolution of composite systems across four generations has helped to improve identification rates substantially. Early research on the holistic composite system EvoFIT, for example, showed that composites created using EvoFIT were named at a higher rate (3.6%) after a two day delay than composites created on the feature-based systems, FACES, PROfit and E-FIT (3.2%, 1.3% and 0.0%, respectively; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005). Considerable research has since been conducted to maximise the naming rates of EvoFIT, with initial advances increasing naming rates to 11% over a two day delay (Frowd et al., 2007; Frowd, Skelton, et al., 2011). More recent research suggests EvoFIT now produces naming rates four times higher than other composite systems, highlighting the extensive advancements that have been made over the past decade (Frowd et al., 2015). Whilst the production of more identifiable images is beneficial for applied settings, by aiding the process of locating suspects, research suggests that the process of composite construction can, in itself, affect an eyewitness' subsequent identification performance. The effects of composite construction are considered further in the next section.

Figure 1.4. Examples of facial composites created of Noel Gallagher using E-FIT, EvoFIT, Photofit, Sketch and Profit (left to right, top to bottom). Taken from Frowd, Carson, Ness, Richardson, et al. (2005)

1.3 Effects of Composite Construction

Whilst composite construction is essential in many criminal investigations to find potential suspects, it is also a concern. Studies of eyewitness memory demonstrate, for example, that the intervening presentation of a potential suspect, between the initial exposure to a target identity and a subsequent line-up, can induce identification errors (Jenkins & Davies, 1985; Kempen & Tredoux, 2012; Topp-Manriquez, McQuiston, & Malpass, 2014). By contrast, the available evidence on whether composite construction exerts similar effects on eyewitness identification accuracy is mixed. Some studies suggest that composite construction is a *helpful* process that enhances eyewitness accuracy. Accordingly, eyewitnesses who create a facial composite of a perpetrator are more likely to follow this up with a correct line-up identification (Davis, Gibson, & Solomon, 2014; Mauldin & Laughery, 1981). Other studies indicate that composite construction can be *harmful*, by impairing eyewitness identification accuracy in a subsequent line-up (Comish, 1987; Kempen & Tredoux, 2012; Topp-Manriquez et al., 2014; Wells, Charman, & Olson, 2005). Additionally, a third strand of research indicates that composite construction incurs *no* effect on subsequent recognition accuracy (Davies, Ellis, & Shepherd, 1978; Davis, Thorniley, Gibson, & Solomon, 2016; Yu & Geiselman, 1993)

1.3.1 The Helpful Effect

Research in support of the helpful effect suggests that the process of composite construction aids an eyewitness' subsequent identification performance. An early study in this field compared the identification accuracy of composite constructors to control participants, who did not construct a composite between the exposure and recognition phases. Using the mechanical composite system, Identi-kit, which requires the selection of individual features (e.g., eyes, nose, mouth) to construct a face, it was found that identification accuracy was higher for composite constructors (90%) than controls (60%; Mauldin & Laughery, 1981).

Although this support for the helpful effect appears promising, there are caveats. For example, whilst Mauldin and Laughery (1981) found composite construction to enhance identification, the performance of composite constructors was improved most when construction occurred immediately prior to recognition. By contrast, identification accuracy was worse when eyewitnesses experienced a two-day interval between composite construction and recognition, suggesting that such delays impede performance. This is an important finding given that this delay could be extended to several weeks in applied settings (Pike & Brace, 2002; Valentine, Pickering, & Darling, 2003).

More recently, Davis et al. (2014) tested the effects of composite construction on identification accuracy using two more sophisticated composite systems than Identi-kit, called E-FIT and EFIT-V. E-FIT is a feature-based software system that requires the recall of individual facial features. EFIT-V, on the other hand, is a holistic composite system that allows eyewitnesses to select images of whole faces from a series of successive arrays. In Davis et al.'s (2014) study, identification accuracy between composite constructors who used E-FIT (63%) and EFIT-V (70%), did not significantly differ. However, the identification rates of both types of composite constructors were higher than the identification rates of controls (45%), who were not required to construct a composite. However, this advantage was obtained with a time delay of only two hours between the initial exposure to a target and subsequent recognition, raising questions of the extent to which this finding generalises to more extended time intervals, that are more relevant to real-life criminal investigations.

To explore this issue, the time delay between initial exposure and composite construction was extended to 30 minutes in a second experiment, whilst the delay between construction and recognition was extended to 72 hours. Under these conditions, the identification accuracy of composite constructors (49%) was, again, higher than the identification accuracy of control participants (35%; Davis et al., 2014). These experiments therefore converge to demonstrate a helpful effect of composite construction on subsequent eyewitness identifications, and suggest that this advantage persists over delays of at least 72 hours between initial exposure to a perpetrator and the final line-up identification (Davis et al., 2014).

Further support for the helpful effect of composite construction comes from a meta-analysis investigating the verbal overshadowing effect (Meissner & Brigham, 2001a). Verbalising information about facial stimuli is thought to create memorial interference and reduce one's facial recognition ability. This detrimental effect is known as the verbal overshadowing effect (Schooler & Engstler-Schooler, 1990) and has been replicated numerous times (for a review see, Meissner & Brigham, 2001a). The recall involved when constructing a feature-based composite, however, is not thought to elicit this effect. In a meta-analysis of eight studies, Meissner and Brigham (2001a), found an improvement in identification accuracy following composite construction, that suggests that composite constructors are 1.56 times more likely to make subsequent correct identifications than control participants (Meissner & Brigham, 2001a). It is possible that the visual recall and visual recognition required during composite construction relies on processes that complement each other (Meissner & Brigham, 2001a). The balance of these processes may therefore mean that, contrary to the verbal overshadowing effect, no memorial interference is caused, resulting in improved identification accuracy.

Overall, these studies demonstrating a helpful effect support the use of composite construction in applied settings and also the use of more recent holistic systems (e.g. Davis et al., 2014). The enhanced identification accuracy rates among composite constructors, comparative to controls, suggests that law enforcement agencies benefit from the use of facial composite systems. However, contradictory evidence also exists to suggest that the process of composite construction should be used cautiously, and indicates that applied settings need to be advised of the potential, harmful consequences for eyewitnesses.

1.3.2 The Harmful Effect

Contrary to the helpful effect, several studies also suggest that composite construction can diminish an eyewitness' subsequent identification performance. Early research in support of such a harmful effect employed no time delay between the initial exposure to a target and composite construction, and a seven-minute delay between construction and recognition (Comish, 1987). This research found that participants who had constructed a facial composite of the target subsequently made more identification errors than control participants, who did not create a composite between the exposure and recognition phases (identification accuracy of 14% and 44%, respectively).

In this instance, Identi-kit was employed and the target face that participants were exposed to prior to composite construction was also an Identi-kit composite. This allowed for an interesting manipulation in the recognition task, which could include foils in the line-up that replicated either a participant's own composite errors or the errors of another participant. This revealed that, when composite constructors viewed a line-up with a foil face that replicated the inaccurate visual features of their own composite, they were more likely to select this modified foil as the target. This suggests that, for composite constructors, the memory for a composite face that they have recently created is more salient than their visual memory of the underlying target. This memory can then bias eyewitness identification in a subsequent line-up.

Several more studies have since replicated this harmful effect of composite construction on eyewitness identification accuracy. Wells et al. (2005), for example, showed participants a target face and then induced a short delay between this exposure and subsequent composite construction, by requiring participants to write down a description of the target. Using the feature-based system FACES, composite constructors then created a facial composite and all participants completed a line-up task after a two-day delay.

Identification accuracy was much higher among control participants (84%) than composite constructors (10%), suggesting, again, that the process of composite construction causes interference with the memory for a target face (Wells et al., 2005).

A more recent replication of Wells et al. (2005) further supports the harmful effect. Again, using the composite system FACES and a two-day delay between composite construction and recognition, identification accuracy was higher for control participants (51%) than composite constructors (23%; Kempen & Tredoux, 2012). Interestingly, Kempen and Tredoux's (2012) findings indicate that mere exposure to facial composites can also generate harmful effects. Identification accuracy rates of composite constructors and participants who viewed, but did not construct a facial composite, did not significantly differ (at 23% and 26%, respectively). These findings therefore indicate that mere exposure to a facial composite may also contaminate one's memory (Kempen & Tredoux, 2012).

The notion that both composite construction and mere exposure to facial composites can elicit harmful effects is further supported by more recent research. In this study, using the composite system FACES, an extended delay of one week was employed between composite construction and recognition (Topp-Manriquez et al., 2014). In this instance, participants who did not construct a composite were also more likely to make a correct identification (31%) than composite constructors (16%; Topp-Manriquez et al., 2014). The notion that mere exposure can impede subsequent identification accuracy is also further supported by this study, as multiple viewings of a facial composite generated harmful effects. Thus, participants who had no exposure to a facial composite during the one week delay were more likely to make an accurate identification (35%) than participants who viewed a composite once (15%) or twice (20%) during this period (Topp-Manriquez et al., 2014). This is a problematic finding given that eyewitnesses are sometimes allowed to retain a copy of their facial composite (McQuiston-Surrett, Topp, & Malpass, 2006).

Overall, support for the harmful effect suggests that composite construction may contribute to memory contamination. Wells et al. (2005) propose that this interference is likely to occur for one of three reasons. Firstly, it is suggested that the process of composite construction *creates* a second memory, the composite memory, which competes with the original memory in any subsequent recognition tasks. Secondly, it is proposed that the original memory is *blended* with the composite memory, carrying forward any potential inaccuracies (see, e.g., Jenkins & Davies, 1985; Neumann, Schweinberger, & Burton, 2013). Thirdly, it is proposed that only the memory for the composite remains intact and this memory *replaces* the original memory altogether (Wells et al., 2005).

The notion that composite constructors retain two memories, both the original memory and the composite memory, as first proposed, fails to explain why witnesses experience a diminished identification performance. That is, if the original memory for the target remained, witnesses should still be able to access this at the time of recognition. The diminished ability of composite constructors to make accurate identifications therefore suggests that the original memory is actually impaired, as opposed to a witness retaining two separate memories (Wells et al., 2005). It is more difficult, at present, to decide between a blending and a replacement account. It is clear, however, that such memory impairments may have serious consequences for applied settings, by exacerbating occurrences of mistaken eyewitness identifications.

1.3.3 The Null Effect

In contrast to evidence for both the helpful and the harmful effect, an additional strand of research exists, which suggests that composite construction elicits no subsequent effect on identification performance. Early support for this effect came from research using the mechanical system Photofit. In this instance, a time delay of either two days or three weeks was employed between target exposure and composite construction, with participants subsequently proceeding immediately onto a recognition task (Davies et al., 1978). Although a slight decrease in recognition accuracy was found for composite constructors, the difference between constructors and controls was not statistically significant (Davies et al., 1978). In line with Davies et al.'s study, a null effect was also found in a study employing no delay between exposure and construction, and a two-day delay between construction and recognition (Yu & Geiselman, 1993). Identi-kit was used in this study and, although constructors (50%) were more cautious in making an identification, rending them more likely than controls (33%) to respond that the target was absent, they were no more likely to make an incorrect identification (Yu & Geiselman, 1993). Finally, more recent work has provided support for the null effect using the holistic composite system EFIT-V. In Davis et al.'s (2016) study, participants viewed a target face and proceeded immediately onto composite construction. Recognition accuracy was then assessed approximately one hour later. In this instance, composite constructors (35%) and controls (32%) were equally able to identify a target identity (Davis et al., 2016).

1.3.4 Limitations of Existing Research

The conflicting findings regarding the effect of composite construction present a dilemma for the justice system. The finding that composite constructors generate the highest amount of incorrect identifications (Comish, 1987; Kempen & Tredoux, 2012; Topp-Manriquez et al., 2014; Wells et al., 2005) is extremely problematic as this infers that law enforcement agencies may be exacerbating mistaken identifications through unsuitable procedures. However, given that there is also support for both a helpful effect (Davis et al., 2014; Mauldin & Laughery, 1981; Meissner & Brigham, 2001a), and a null effect (Davies et al., 1978; Davis et al., 2016; Yu & Geiselman, 1993), the justice system cannot be accurately advised on best-practice until a scientific consensus is reached.

Moreover, the considerable applied importance of the facial composite literature means that research must replicate real-life police procedures as closely as possible. In this context, the lack of forensically relevant time delays within existing research is concerning, as it is not yet clear how applicable the available scientific data are to applied settings. In these settings, composite construction is usually attempted within 24-36 hours of witnessing a crime (Frowd, Carson, Ness, Richardson, et al., 2005). Eyewitnesses may then experience a delay of several weeks before being required to identify the perpetrator from a suspect line-up (Pike & Brace, 2002; Valentine et al., 2003). The short time delays that are employed in the scientific research to date, ranging from no delay (e.g., Mauldin & Laughery, 1981) to a maximum delay of one week (Topp-Manriquez et al., 2014), therefore raise questions of the extent to which these findings represent effects that are likely in applied settings.

Additionally, much of the available research uses feature-based composite systems. Most notably, all of the existing support for a harmful effect has employed such systems (e.g. Topp-Manriquez et al., 2014; Wells et al., 2005). Considering that holistic systems are increasingly becoming the standard (Frowd et al., 2012; Solomon et al., 2012), and replacing feature-based systems, it currently remains unclear which of the documented effects extends to the holistic systems that are currently in police use.

1.4 Morphing Composites

The low naming rates that composite construction often generates suggests that facial composites contain inaccuracies that compromise their successful identification. In instances where multiple witnesses create a composite of the same target identity, however, averaging these individual composites together can produce a better overall likeness to the target face

(Hasel & Wells, 2007; Valentine et al., 2010). The process of averaging individual composites together is referred to as morphing. Morphing improves composite quality by averaging out any uncorrelated errors but retaining features present in every individual composite (Hasel & Wells, 2007; Valentine et al., 2010). A clear advantage for morphing has been found. Morphed composites are rated as more similar to the target than individual composites (Hasel & Wells, 2007), and produce higher naming rates comparative to individual composites (Valentine et al., 2010).

Evidently, morphing composites constructed by *different* witnesses (between-witness morphs) generate the most accurate resemblance to the original target face, producing naming rates of 44% (Valentine et al., 2010). In comparison, morphing composites constructed by the *same* witness (within-witness morphs) generate naming rates of 32%, whilst individual composites are less recognisable and named at 20% (Valentine et al., 2010).

The advantageous effect of morphing, and the advantage of between-witness morphs, shows that there must be variation in composite quality between witnesses. This variation has been accepted within the composite literature without the role of individual differences being researched. People are known to vary in their abilities to perceive and recognise faces (Bindemann, Avetisyan, & Rakow, 2012; Bindemann, Brown, Koyas, & Russ, 2012; Burton, White, & McNeill, 2010; Russell, Duchaine, & Nakayama, 2009), meaning that not all eyewitnesses may be equally equipped for the task of composite construction. In the next section, these individual differences are considered in more detail.

Figure 1.5. Taken from Valentine et al (2010). An example of individual composites created of the actor, James Alexandrou, who plays Martin Fowler in the soap opera Eastenders. All four composites were created by different witnesses and highlight the variation of composite quality between witnesses.

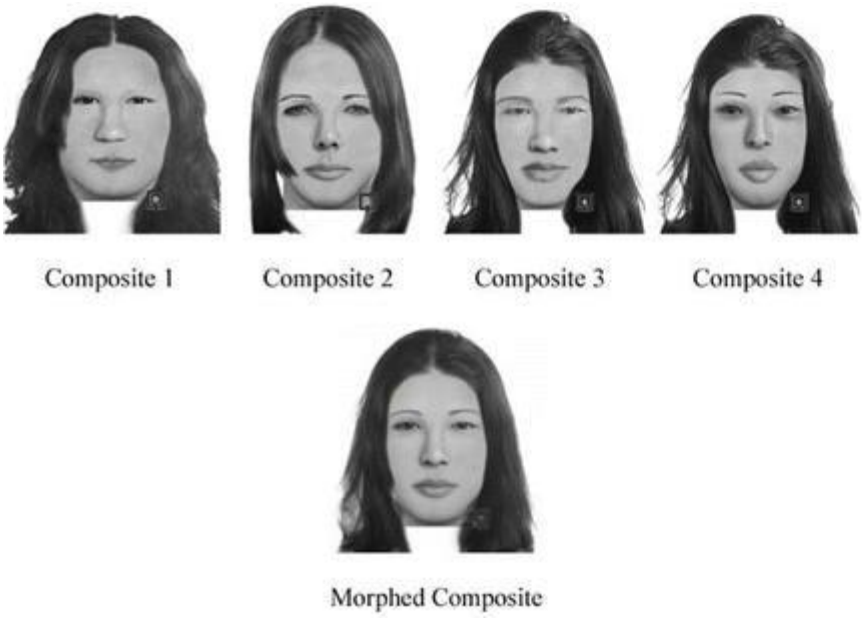

Figure 1.6. Taken from Hasel and Wells (2007). An example of four individual composites and a between witness morph. Note the differences between the four individual composites.

1.5 Individual Differences

The face recognition literature demonstrates enormous variation between people's abilities to recognise both familiar (Duchaine, Germine, & Nakayama, 2007; Robertson et al., 2016) and unfamiliar faces (Bindemann, Avetisyan, et al., 2012; Bobak, Dowsett, & Bate, 2016; Burton et al., 2010; Russell et al., 2009). The range of recognition abilities is evident on, for example, unfamiliar face matching tasks whereby subjects are required to indicate whether a face pair is the same person (match) or different people (mismatch). Individual differences on such tasks result in matching accuracy ranging from near chance (51%) to ceiling (100%; Burton et al., 2010). Likewise, on familiar face matching tasks using celebrity identities, matching accuracy can range from 73% to 100% (Robertson et al., 2016). The discrepancy in recognition abilities is also evident on standardised tests of unfamiliar face recognition whereby subjects are required to identify a previously seen face on test items consisting of both the target and distractor faces. On these tests, performance between individuals can vary from 50% (Duchaine & Nakayama, 2006) to 100% (Russell et al., 2009). These findings therefore converge to demonstrate a broad range of recognition abilities across individuals.

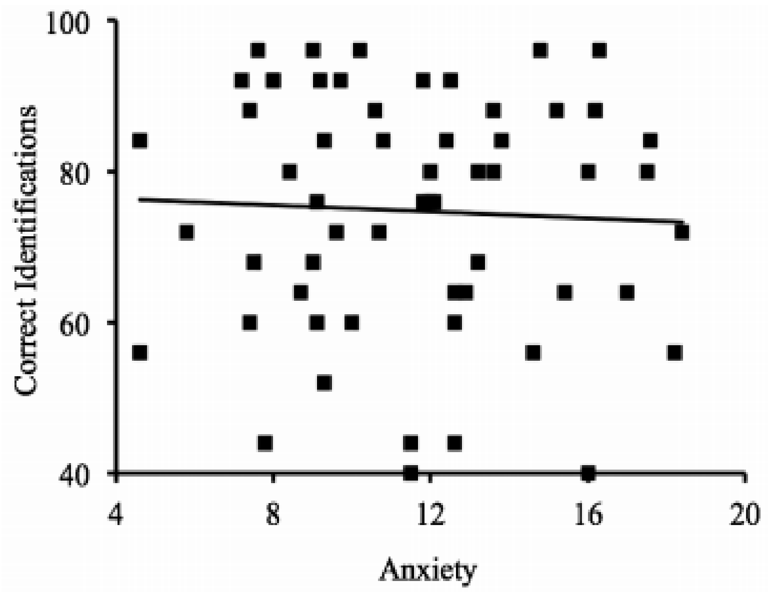

Figure 1.7. Taken from Megreya and Bindemann (2013). An example of the broad distribution of individual differences on a face matching task.

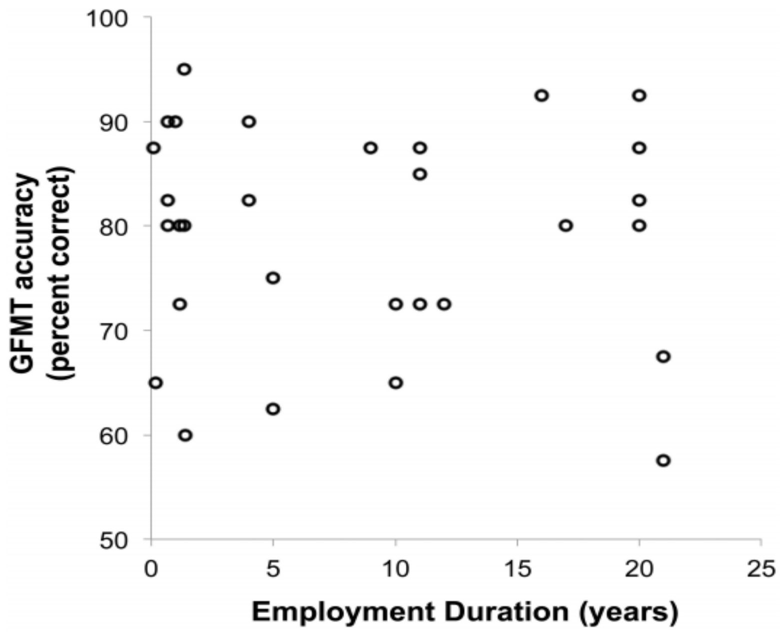

Figure 1.8. Taken from White, Kemp, Jenkins, Matheson, and Burton (2014). A second example of large individual differences on a matching task.

Despite the broad distribution of abilities between individuals, an individual's face recognition skill appears to remain relatively stable across different tasks, suggesting that performance on one task can be indicative of performance on another (Bobak et al., 2016; Robertson et al., 2016; Russell et al., 2009). For example, individuals who excel at face matching tasks are able to replicate high performance levels across different tests of face matching (Bobak et al., 2016), including familiar and unfamiliar faces (Robertson et al., 2016). Likewise, individuals who perform poorly on unfamiliar face recognition tests also perform poorly on tests of familiar face recognition (Duchaine et al., 2007). In addition, face recognition skills remain relatively stable across time, with training for face recognition bearing little impact on accuracy (White et al., 2014; Wilmer, 2017). However, the stability of these skills is also subject to individual differences, with consistency levels varying between individuals (Bindemann, Avetisyan, et al., 2012). Despite the enormous range of recognition abilities, face recognition appears to be hardwired and have a genetic basis. This heritability is evident in the recognition performance of twins. The scores of monozygotic twins on face recognition tasks are more highly correlated than that of dizygotic twins, providing support for the notion that genetic differences contribute to the broad distribution of abilities (McKone & Palermo, 2010; Wilmer et al., 2010). If face identification ability is relatively stable, then a measure of this may also provide insight into a person's ability to produce a good-quality facial composite.

1.5.1 Individual Differences and Eyewitness Accuracy

Although face recognition has been studied theoretically, the discovery that face recognition skills span a broad distribution also has applied importance. An individual's ability to perform on standardised face recognition tests may relate to that individual's accuracy as an eyewitness. This notion was investigated in a recent study which presented participants with two tasks: an eyewitness test and a standardised face test (Bindemann, Brown, et al., 2012). In the eyewitness test, participants viewed a video of a staged crime and were subsequently required to identify the perpetrator from a ten-face line-up. This test was followed by a face test, comprised of forty trials of the 1-in-10 task (Bruce et al., 1999). The 1-in-10 task requires participants to study a single target face and then indicate whether the target is present in a subsequent line-up (see Figure 1.9). Performance on the two tests were positively correlated, suggesting that an individual's ability to remember and recognise faces relates to their aptitude as an eyewitness (Bindemann, Brown, et al., 2012). Moreover, individuals who excelled on the face test, scoring 90% or higher, were always accurate in their eyewitness identification (see Figure 1.10).

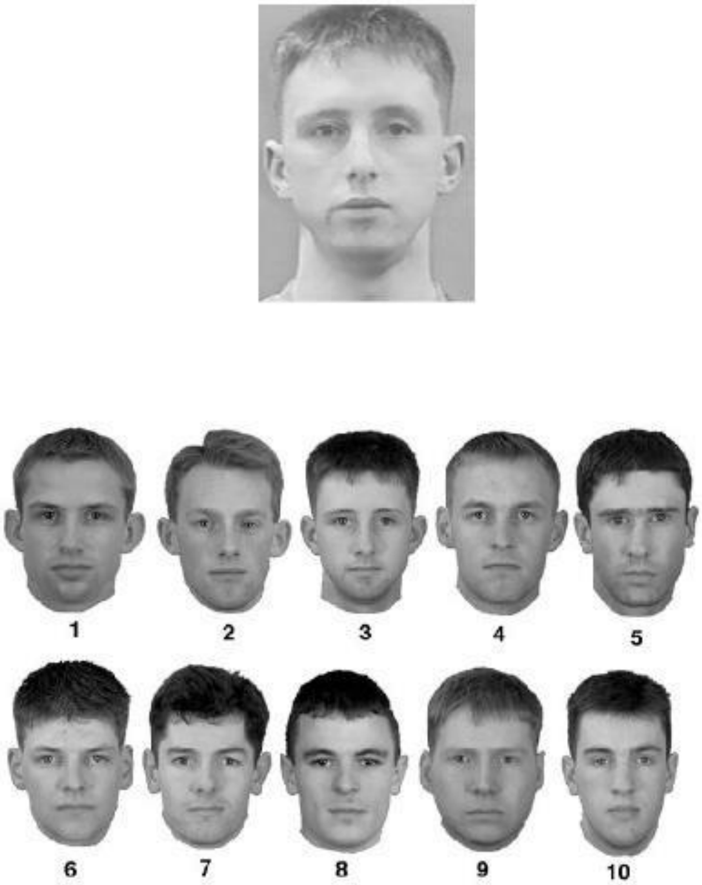

Figure 1.9. Taken from Bruce et al. (1999). An example trial of the 1-in-10 task. Participants are required to either select a face from the ten-face line-up that matches the target identity, or indicate that the target is absent.

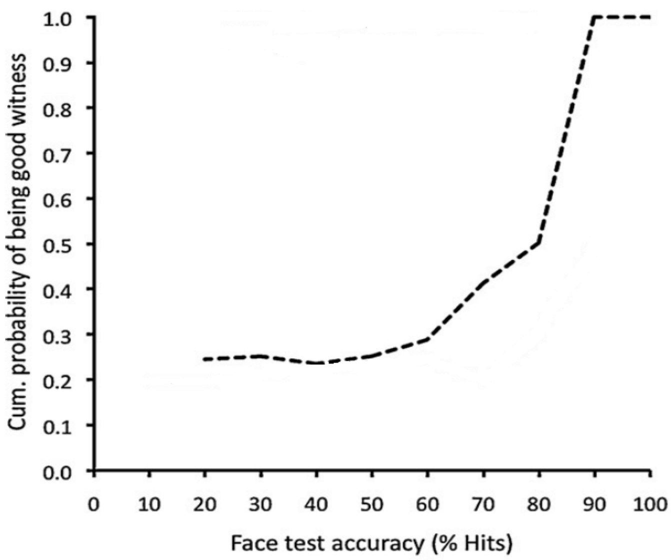

Figure 1.10. Taken from Bindemann, Brown, et al. (2012). The correlation between scores on the 1-in-10 task and accurate identifications on the eyewitness test.

The relationship between an individual's performance on a standardised face test and their eyewitness accuracy has also been found in research involving heightened levels of personal, emotional distress. During training, US military personnel complete an experiential phase of training, which involves recruits being captured, placed in a mock prisoner of war camp (POWC) and being subjected to interrogation by an unknown instructor (Morgan et al., 2000, 2007). Morgan et al. (2007) investigated the eyewitness accuracy of military personnel under these stressful circumstances by asking them to identify their interrogator from ten sequentially presented photographs. Prior to the eyewitness test, recruits also completed the Weschler Face Test (1997), which required them to study 24 target faces and subsequently identify each target face from 48 sequentially presented photographs. In this instance, although more than one in three recruits were unable to accurately identify their interrogator, performance on the face test was significantly associated with eyewitness identification accuracy (Morgan et al., 2007). These findings therefore support the notion that performance on standardised face recognition tests relates to the probability of being a good eyewitness but, most importantly, indicate that this relationship persists outside of laboratory settings, in distressing circumstances.

Evidently, there are substantial individual differences in face recognition and this broad distribution has considerable applied importance, relating to an individual's aptitude as an eyewitness (Bindemann, Brown, et al., 2012; Morgan et al., 2007). What remains unclear, however, is whether these individual differences correspond to other aspects of eyewitness testimony, aside from a witness' final identification accuracy. As aforementioned, holistic composite systems align composite construction more closely with recognition (Frowd, 2017). Based on this improved theoretical approach to composite construction, an individual's ability to perform on standardised face recognition tests may also correspond to their ability to construct an accurate facial composite of a target face. So far, however, no direct evidence exists to address this question. Investigating this relationship may therefore provide a means of improving composite identification rates, by selecting only the most able eyewitnesses for the task, likely to construct the most recognisable composite. In the next section, another means of improving composite identification rates is explored through the provision of contextual information.

1.6 Context and Recognition

In police investigations, when facial composites are released to the media, the composite image is displayed with context (for examples see Crimestoppers, 2018; Metropolitan Police, 2016). The context may be explicit, providing details of the perpetrator's appearance (e.g. age, height, clothing), or implicit, with the image being displayed in locations which the offender is known to frequent. The composite literature indirectly acknowledges the worth of context through paradigms that may facilitate recognition. Such paradigms include, for example, constructing facial composites of unknown professional snooker players to be subsequently named by snooker fans (Frowd, Pitchford, et al., 2011). Likewise, other studies constrain context by providing additional information about the target, such as their occupation, during naming tasks (Valentine et al., 2010). Restricting the range of possible identities in this manner is known as cued naming and produces higher naming rates than spontaneous naming, whereby no additional information is provided (see, e.g., Frowd et al., 2007; Frowd, Carson, Ness, Richardson, et al., 2005; Valentine et al., 2010). The increase in naming rates associated with cued naming suggests that any contextual information aids the recognition of facial composites. However, the scale of this difference is currently unclear as the effect of context on composite recognition has not been investigated systematically.

Whilst the composite literature has not yet investigated the effect of context, other domains show how context can facilitate recognition. Semantic priming refers to the notion that response times to targets are faster if the target is preceded by semantically related primes (Seidenberg, Waters, Sanders, & Langer, 1984). Semantic priming has shown that familiar face recognition is affected by the amount of semantic information known, with faces being recognised more quickly when more semantic information is provided (Bruce, Dench, & Burton, 1993; Burton, Bruce, & Johnston, 1990). Whereas these effects are typically observed via the speed of participants' responses during face classification, the explicit recognition of unfamiliar faces has also been found to be affected by the amount of context provided. Kerr and Winograd (1982) exposed participants to unfamiliar faces that were presented with either zero, one, two or three descriptive phrases relating to the target's personality, occupation and hobbies. A significant difference was found between the zero and one-phrase conditions, suggesting that encoding at least one descriptive phrase about a face aids subsequent recognition. The similar recognition accuracy across the one, two and three-phrase conditions, however, suggests an all-or-none response to context, rather than a gradual effect on recognition based on the amount of context provided (Kerr & Winograd, 1982).

To explore this issue further, Kerr and Winograd (1982) conducted an additional experiment that exposed participants to the descriptive phrases at both the encoding and recognition phases. In this instance, a significant effect of elaboration was found, suggesting again that context does facilitate recognition, but this effect was more profound when contextual information was reinstated at the time of recognition.

	Elaboration (Number of Phrases)		
Test Condition	0	1	2
No Context	.82	.85	.88
Context Present	.82	.93	.92

Figure 1.11. Taken from Kerr and Winograd (1982). Context reinstatement facilitated recognition in both the one and two-phrase conditions.

The reinstatement of context has also been found to facilitate recognition in field experiments investigating eyewitness accuracy. Kraflca and Penrod (1985) asked shop assistants to identify a customer from a previous encounter that had taken place either two hours or 24 hours earlier. Contextual information was reinstated by providing a physical cue (a non-photographic form of the customer's identification), and a mental cue, by asking the assistant to recall the interaction with the customer. In this instance, context reinstatement significantly increased the ability of an eyewitness to make an accurate identification of the customer from a photographic line-up (Krafka & Penrod, 1985).

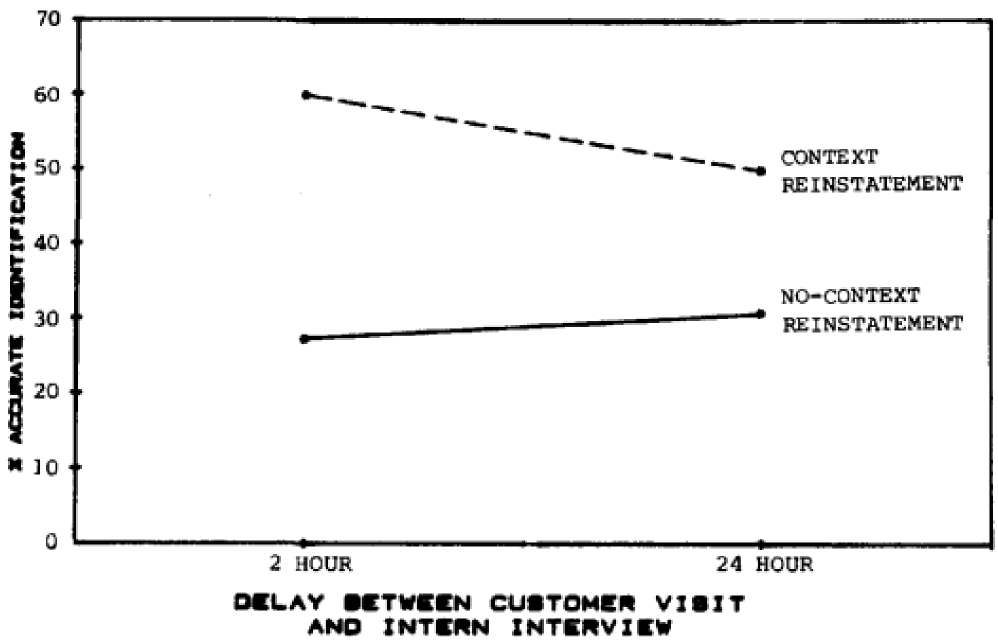

Figure 1.12. Taken from Kraflca and Penrod (1985). The percentage of correct eyewitness identifications in target-present photographic line-ups.

These findings all support the notion that individuals are more likely to recognise a face if additional information is provided about the target (Burton et al., 1990; Kerr & Winograd, 1982; Krafka & Penrod, 1985). Moreover, recognition is most enhanced when context is reinstated at the time of recognition (Kerr & Winograd, 1982; Krafka & Penrod, 1985), meaning that these findings may have important direct implications for composite naming rates. Given that composites are displayed with context, investigating the effect of context on composite naming has useful implications for criminal investigations.

1.7. The Structure of this Thesis

The purpose of this thesis is to investigate the relationship between facial composites and person identifications. The first empirical chapter addresses the debate regarding the effect of composite construction, whilst recognising the need for research settings to employ forensically relevant delays. To establish whether either the helpful, the harmful, or the null effect of composite construction persists over forensically relevant delays, Experiment 1 applied prolonged intervals between exposure, composite construction and recognition phases. A time delay of 24 to 48 hours was employed between the initial exposure to a target face and composite construction. Recognition performance was subsequently tested using a line-up task ten weeks later. The delays used in Experiment 1 therefore align more closely with applied settings. In addition, a second time condition was employed, which reflects the short delays used in existing research (e.g. Davis et al., 2014; Kempen & Tredoux, 2012; Mauldin & Laughery, 1981; Wells et al., 2006). In this time condition, participants proceeded immediately onto composite construction after viewing a target face, and completed a line-up task after a two-day delay. The comparison of the recognition accuracy across these two time conditions therefore allowed for an interesting investigation into the effect of composite construction that is most likely to extend to applied settings.

Chapter 3 builds on the notion that an individual's performance on standardised recognition tests corresponds to their eyewitness identification accuracy (Bindemann, Brown, et al., 2012; Morgan et al., 2007). As composite construction is only a necessary task when the perpetrator is unknown, Experiment 2 investigates whether an individual's performance on tests of unfamiliar face recognition also correlates to their ability to produce accurate facial composites. This was achieved using two standardised tests of recognition: The Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006) and the 1-in-10 task (Bruce et al., 1999). Facial composites were then created on a holistic composite system that relies on face recognition (Frowd, 2017), to explore whether individual differences in the ability to remember and recognise unfamiliar faces can provide an index of which eyewitnesses are equipped best for the task of composite construction.

The final empirical chapter investigates the effect of context on composite naming rates. Facial composites of celebrity identities were presented to participants with varying levels of context. Composites were either displayed with no context or were accompanied by information ranging from the target's occupation (minimal context) to the target's age, occupation and what they were best known for (maximum context). The relationship between the amount of context provided and naming rates was investigated here to provide a useful indicator to applied settings regarding the amount of information required to facilitate recognition.

Chapter 2

The Effect of Composite Construction on Line-up Performance Using Forensically Relevant Time Delays

Introduction

In the previous chapter, the difficulty of composite construction was evident through factors such as low naming rates (e.g. Frowd et al., 2007; Valentine et al., 2010). In recent years however, intensive developments have been made to improve the effectiveness of holistic systems such as EvoFIT, which have resulted in increased identification rates (Frowd et al., 2012; Frowd et al., 2013). Whilst the production of more identifiable images may aid the process of locating and apprehending suspects, whether composite construction, in itself, can exert an effect on an eyewitness' subsequent identification performance currently remains unresolved.

A number of conflicting findings exist, which indicate that the process of composite construction can produce a number of different effects on an eyewitness' subsequent ability to accurately identify a target face. On one hand, research indicates that the process of composite construction can produce a *helpful* effect, whereby composite constructors are *more* likely to make subsequent correct identifications (Davis et al., 2014; Mauldin & Laughery, 1981). However, a *harmful* effect has also been reported, such that constructors are *less* likely to make accurate identifications (Comish, 1987; Kempen & Tredoux, 2012; Topp-Manriquez et al., 2014; Wells et al., 2005). In addition, there is also evidence to suggest that composite construction does neither positively or negatively influence the subsequent recognition of a target (Davies et al., 1978; Davis et al., 2016; Yu & Geiselman, 1993).

Early research in this domain provided support for the helpful effect by demonstrating that identification accuracy was higher for composite constructors (90%), compared with control participants (60%), who had not been required to construct a facial composite (Mauldin & Laughery, 1981). This early study utilised the mechanical system Identi-kit, which requires composite constructors to select individual facial features (e.g., eyes, nose) that are then superimposed to create a composite image. In this instance, participants viewed a target face and either proceeded immediately onto composite construction and a recognition task, or experienced a two-day delay between the initial exposure to the target and subsequent construction and recognition tasks. Although it was found that composite construction facilitated recognition performance, this effect was greater when composite construction occurred immediately prior to recognition. The facilitating effect was less profound when participants experienced a two-day delay between composite construction and the recognition task (Mauldin & Laughery, 1981).

Despite this, more recent research has also provided support for the helpful effect of composite construction on subsequent recognition performance. Davis et al. (2014), for example, employed two different composite systems, E-FIT and EFIT-V, to test the effect of composite construction on subsequent identification accuracy. The two composite systems utilised in this study differ substantially to one another. Whilst E-FIT is a system that requires constructors to recall individual facial features, EFIT-V is a more recent holistic system that allows constructors to view and select images from whole-face arrays. In this study, participants were exposed to a target face before proceeding immediately onto composite construction and then to a recognition task. The identification accuracy of constructors using both E-FIT (63%) and EFIT-V (70%), was significantly higher than control participants (45%), who were not required to construct facial composites. The helpful effect of composite construction on subsequent identification accuracy was also replicated in a second experiment employing longer time intervals between tasks. In this experiment, Davis et al. (2014) utilised a time delay of 30 minutes between a participant's initial exposure to a target face and composite construction, and a delay of 72 hours between construction and recognition. Once again, composite constructors (49%) outperformed control participants (35%) in their ability to make a subsequent accurate identification of the original target (Davis et al., 2014). The findings of Mauldin and Laughery (1981) and Davis et al. (2014) therefore converge to demonstrate a helpful effect of composite construction on subsequent recognition accuracy, despite key differences in methodology (e.g., type of composite system used).

In contrast to the findings discussed above, several studies have documented a harmful effect, suggesting that composite construction can hinder subsequent recognition accuracy. Comish (1987), for example, exposed participants to a target face before requiring them to proceed immediately onto composite construction, using the composite system Identi-kit. A seven-minute delay was then employed between composite construction and recognition. In this instance, the recognition performance of composite constructors (14%) was significantly poorer than that of control participants, who did not undergo the task of composite construction (44%; Comish, 1987).

The harmful effect has since been replicated several more times. In a more recent study, Wells et al. (2005) employed a short time delay between initial exposure to a target identity and composite construction, by requiring participants to write down a description of the target. Facial composites were then created on the feature-based system FACES, and recognition accuracy was assessed after a two-day delay. In support of the harmful effect, recognition performance was, again, higher among control participants (84%) than composite constructors (10%; Wells et al., 2005). Wells et al.'s (2005) study has since been replicated and similar findings are evident. In their replication of Wells et al. (2005), Kempen and Tredoux (2012) employed a ten-minute delay between target exposure and composite construction, and a two-day delay between construction and recognition. Again, the feature-based composite system FACES was employed, and control participants made more correct identifications (51%) than composite constructors (23%; Kempen & Tredoux, 2012).

Finally, additional support for the harmful effect is provided by recent research employing a longer time interval between composite construction and recognition (Topp-

Manriquez et al., 2014). In this instance, participants experienced a five-minute delay between viewing a target face and composite construction, with recognition performance being assessed after a delay of one week. Recognition accuracy was, again, higher for controls (31%) than constructors (16%; Topp-Manriquez et al., 2014). These experiments are therefore unified in their findings of a harmful effect, suggesting that composite construction may not be a suitable procedure if the same eyewitness is subsequently required to identify the original perpetrator (Comish, 1987; Kempen & Tredoux, 2012; Topp-Manriquez et al., 2014; Wells et al., 2005).

In addition to evidence suggesting the existence of both a helpful and a harmful effect, a third finding exists, which suggests that composite construction incurs *no* effect on subsequent recognition accuracy. Early research using the mechanical composite system Photofit, employed a delay of either two days or three weeks between target exposure and composite construction (Davies et al., 1978). Participants were subsequently required to proceed immediately onto a recognition task. Although it was found that composite constructors tended to make more errors than controls, this finding was not statistically significant (Davies et al., 1978). Likewise, a null effect was also found in a study employing no delay between exposure and construction, and a two-day delay between construction and recognition (Yu & Geiselman, 1993). This study made use of the composite system Identi-kit and, although composite constructors (50%) were more likely than controls (33%) to incorrectly respond that the target was absent, constructors were no more likely to make an incorrect identification (Yu & Geiselman, 1993). Finally, research employing the more recent holistic system EFIT-V found that composite constructors (35%) and controls (32%) were equally able to identify a target identity (Davis et al., 2016). In this instance, participants proceeded immediately onto composite construction after viewing a target face, and completed a recognition task approximately one hour after initial exposure (Davis et al., 2016).

The contrasting findings within existing research mean that it is currently unclear whether the task of composite construction is a suitable procedure for eyewitnesses. Considering the applied importance of this domain, it is therefore necessary to reach a scientific consensus that is able to accurately advise law enforcement agencies on best-practice. Moreover, large discrepancies are evident between the time intervals used in applied and research settings. These discrepancies make it difficult to determine which of the documented effects apply to forensic practice. Most notably, the majority of existing research has employed a very short time interval between a participant viewing a target face and the subsequent task of composite construction (see Table 2.1). Likewise, the time intervals employed between composite construction and recognition are not reflective of forensic practice, and range from no delay (e.g. Mauldin & Laughery, 1981), to a maximum delay of one week (Topp-Manriquez et al., 2014). In contrast to the delays employed within research settings, eyewitnesses in criminal investigations experience much longer delays. For example, a two-day delay is typically experienced between witnessing a crime and composite construction (Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Frowd, Carson, Ness, Richardson, et al., 2005). Eyewitnesses may then also experience a delay of several weeks before being required to identify the original perpetrator from a suspect line-up (Pike & Brace, 2002; Valentine et al., 2003).

Table 2.1. A summary of empirical studies investigating the effect of composite construction on identification performance. Note that Delay 1 corresponds to the delay between initial exposure and composite construction. Delay 2 refers to the delay employed between construction and recognition.

	Delay 1	**Delay 2**	**Composite system**	**Effect**
Mauldin & Laughery (1981)	Immediate or 2 day	Immediate or 2 day	Identi-kit	Helpful
Davis, Gibson, & Solomon (2014)	Immediate (Exp 1)	Immediate (Exp 1)	E-FIT & EFIT-V	Helpful
	30 Minutes (Exp 2)	72 hours (Exp 2)		
Comish (1987)	Immediate	7 minutes	Identi-kit	Harmful
Wells, Charman, & Olson (2005)	After writing description (Exp 1 & 2)	2 day (Exp 1 & 2)	FACES	Harmful
Kempen & Tredous (2012)	10 minutes	2 day	FACES	Harmful
Topp-Manriquez, McQuiston, & Malpass (2014)	5 minutes	1 week	FACES	Harmful
Davies, Ellis, & Shepherd (1978)	2 day or 3 weeks	Immediate	Photofit	Null
Davis, Thorniley, Gibson, & Solomon (2016)	Immediate	1 hour	EFIT-V	Null
Yu & Geiselman (1993)	Immediate	2 day	Identi-kit	Null

In addition, although existing research varies in their choice of composite system, many studies employ feature-based systems (see Table 2.1). Interestingly, all research documenting a harmful effect employed such systems (e.g., FACES). Considering that feature-based systems produce composites that are often difficult to recognise (see, e.g. Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005), it is possible that composite constructors' memories are adversely impacted because of the poor likeness shared between target and composite (see, e.g., Davis et al., 2016). Considering that more recent holistic systems are thought to provide a better interface to memory, and have resulted in higher identification rates (Frowd, Pitchford, et al., 2012; Solomon et al., 2012), it is therefore possible that such adverse effects would not extend to holistic systems.

The current study therefore aimed to test the effect of composite construction on subsequent identification performance using forensically relevant delays and the holistic composite system EvoFIT. The recognition accuracy of composite constructors and control participants was tested either two days or ten weeks after initial exposure to a target identity. In the *two-day* condition, participants viewed a target face and proceeded immediately onto a cognitive interview and composite construction. Recognition accuracy was then assessed two days later. In the *ten-week* condition, participants viewed a target face and then completed a cognitive interview and composite construction 24-48 hours later. Recognition accuracy was then assessed ten weeks later. This design allowed for a condition that reflects the delays typically used in research settings (two-day condition), and a comparison group which experienced forensically relevant delays (ten-week condition). This experiment therefore provides an indication of whether the effects elicited in research settings extend to applied settings, which will, in turn, help to determine whether composite construction is a suitable task. Considering the contradictory findings in the literature, as well as the lack of forensically relevant delays in existing research, *a priori* predictions regarding the effect of composite construction are difficult. It is expected, however, that time elapsed prior to identification will affect recognition accuracy. Thus, recognition performance of participants in the two-day condition is expected to be higher than of those in the ten-week condition.

Method

Participants and design

Eighty students (61 female, 19 male) from the University of Kent, with a mean age of 24 years (*SD* = 7.7; range 18-57), participated in this experiment in return for either course credit or a fee of £10. Only Caucasian participants were eligible for this study. This eligibility requirement was enforced because the stimuli comprised of Caucasian target faces, and the wider face recognition literature provides evidence of an own-race bias (for reviews see, Meissner & Brigham, 2001b; Sporer, 2001). This experiment was a 2 (composite condition: composite constructors vs. no construction) x 2 (time delay: two days vs. ten weeks) between-subjects design with *N* = 20 per group.

Stimuli and procedure

Part 1. Video exposure to a target face. Eight Caucasian males were selected as target identities from the Glasgow Unfamiliar Face Database (GUFD; Burton et al., 2010). A short video clip of each target identity was extracted from the GUFD, which depicted each target performing a sequence of tasks including, looking from left to right, up and down, and reciting letters of the alphabet. The average length of video was 51 seconds. Each video was linked to an online Qualtrics survey to ensure participants could only access their allocated video clip once.

All participants firstly entered their demographic details into the Qualtrics survey before proceeding to watch one of the eight target videos. The order of presentation for target identities was automatically randomised after each participant. Participants in the *two-day* condition proceeded directly onto part 2 after viewing the video clip. For participants in the *ten-week* condition, part 1 of the study ended after the video clip was complete and participants were informed that they needed to complete part 2 within 24-48 hours.

Part 2. Cognitive interviews and composite construction. Participants in the two-day condition proceeded directly to this stage of the study, which is in accordance with existing research that has employed either short, or no time delays between exposure to a target face and composite construction (e.g. Comish, 1987; Davis et al., 2014; Wells et al., 2005). Conversely, participants in the ten-week condition experienced a 24-48 hour time delay between part 1 and part 2, which is more closely aligned with forensic practice (Frowd, Carson, Ness, Richardson, et al., 2005).

In this stage of the experiment, all participants were subjected to a cognitive interview. The cognitive interview began with free recall, whereby participants described the target face in an uninterrupted format. Participants described the target's features in the order of their choice, with no time pressure. A cued recall phase then began whereby the researcher repeated the details that the participant had provided, and encouraged further recall of each facial feature. The cognitive interview is a technique used elsewhere in the composite literature (Frowd et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005), and is known to assist recall (for a review see, Memon, Meissner, & Fraser, 2010). Following the cognitive interview, participants were assigned to either a composite construction condition or a control condition.

Participants in the composite construction condition were instructed that they would create a facial composite of the target face they had previously seen in the video clip and were introduced to the software system EvoFIT, version 1.6. EvoFIT was selected for use in the current study as it has been shown to produce more identifiable images (Frowd, 2012; Frowd, 2017; Frowd, Skelton, Atherton, & Hancock, 2012), which elicit higher correct naming following a forensically relevant delay compared to other systems (Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005). The process requires composite constructors to view a number of screens depicting whole-face arrays and repeatedly select the most recognisable image. This procedure is thought to align composite construction more closely with the natural, holistic processes involved in face recognition, therefore relying less on recall (Frowd, 2017), which is known to be a more difficult task (Davis et al., 1961; Postman et al., 1948). Composite constructors first make selections based on facial shape, and select the images that they feel most accurately represent the size and placement of the target's facial features. The focus is then shifted to facial texture, with composite constructors selecting the images that are most recognisable, based on differences between the greyscale colouring of features such as the target's eyes, eyebrows and skin tone. Selections are then 'bred' together to generate a facial composite (for a review see, Frowd, 2017). Characteristics such as age, weight and trustworthiness can subsequently be manipulated using software tools, to improve the overall likeness to the target face. The researcher acted as an operator for the system for all composite constructors. For composite constructors, part 2 took approximately 45 minutes and participants were dismissed following completion of their facial composite. Control participants were *not* required to construct a facial composite and were therefore dismissed after the cognitive interview was complete.

Figure 2.1. An example of a target identity shown during a video clip (top) and facial composites created of the target in both the two-day (left) and ten-week (right) time conditions.

Figure 2.2. A second example of a target identity (top) and facial composites created of the target in the two-day (left) and ten-week (right) time conditions.

Part 3. Line-up identifications. After a delay of either two days or ten weeks, participants' identification accuracy was assessed on a target-present line-up task (see Figure 2.3). All participants were sent an email containing an eight-face line-up that contained the target face they had originally viewed in the video clip. Participants were instructed to study the line-up and were asked to identify the original target by indicating the number corresponding to its position in the line-up. Participants were informed that the target "may or may not be present". Confidence ratings were also obtained by asking participants to indicate how confident they were that they had made an accurate identification on a scale ranging from one ("not at all confident") to seven ("very confident").

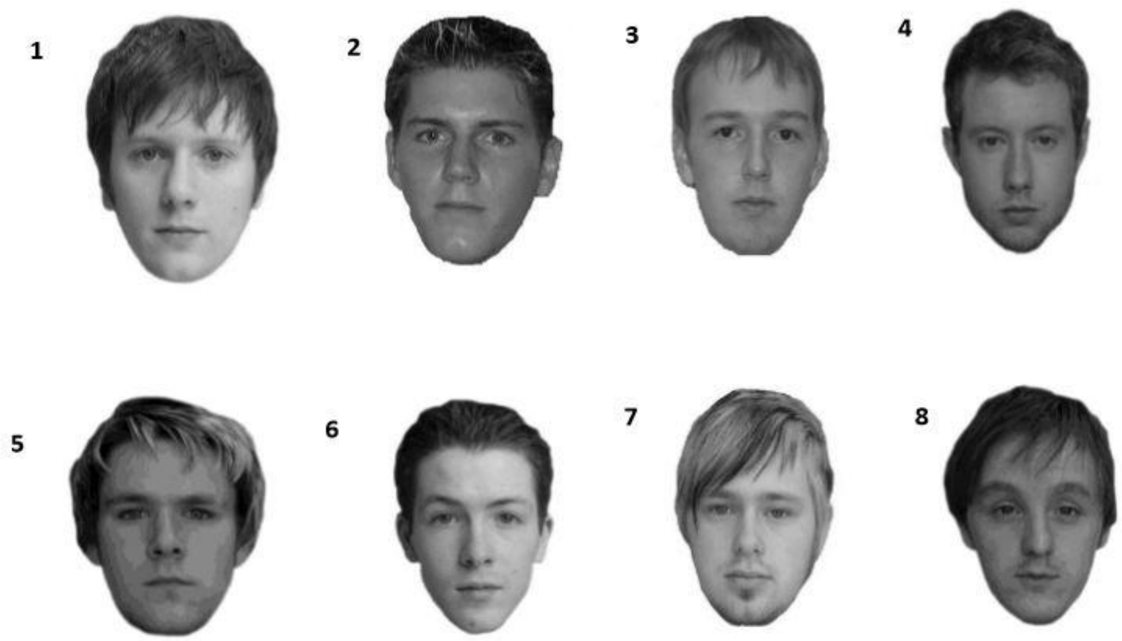

Figure 2.3. An example of a target-present line-up task.

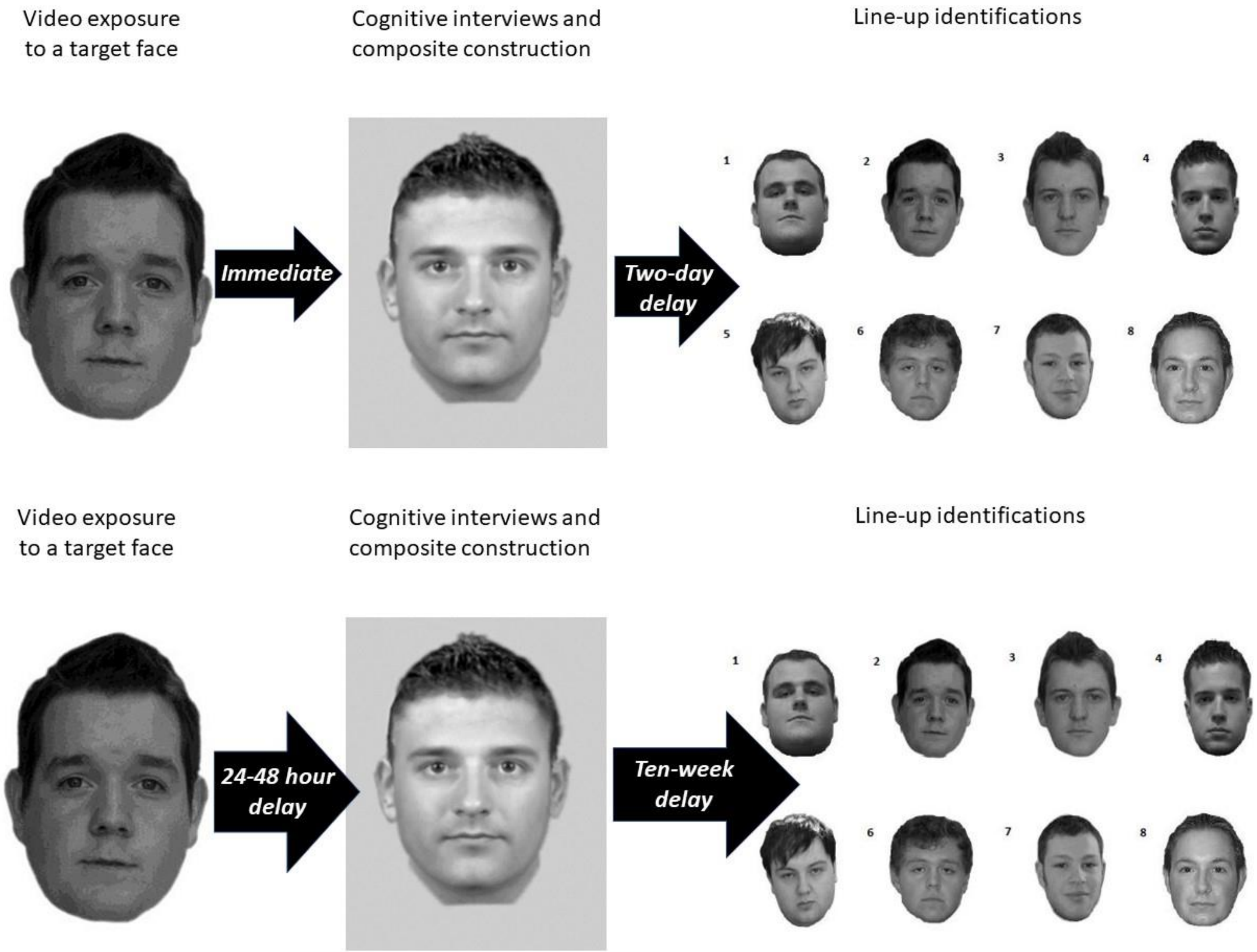

Figure 2.4. Examples of the time delays used during the two-day (top) and ten-week (bottom) time conditions.

Results

Recognition accuracy

Recognition accuracy was compared across both the composite conditions and time intervals. These data are provided in Figure 2.5. Overall, 17 participants made incorrect identifications, whilst 63 participants correctly identified the target. However, there was a clear difference in recognition accuracy across time intervals. In the two-day condition, only two participants made incorrect identifications. In contrast, 15 participants made incorrect identifications in the ten-week condition. The data was subsequently broken down into hits (reflecting a correct identification of the target), misses (incorrect responses that the target is absent) and misidentifications (erroneous identifications of another line-up face as the target). Overall, recognition accuracy was high, with participants recording 79% of hits, 16% of misidentifications and 5% of misses. A 2 (composite condition: composite constructors vs. no construction) x 2 (time delay: two days vs. ten weeks) chi-square test was then conducted. The chi-square test revealed that recognition accuracy was not significantly affected by composite condition, $\chi2$ (1, n = 80) = 0.67, p = .41, but was affected by time delay, $\chi2$ (1, n = 80) = 12.62, p = .001. Recognition accuracy was not affected by target identity, $\chi2$ (7, n = 80) = 8.78, p = .27.

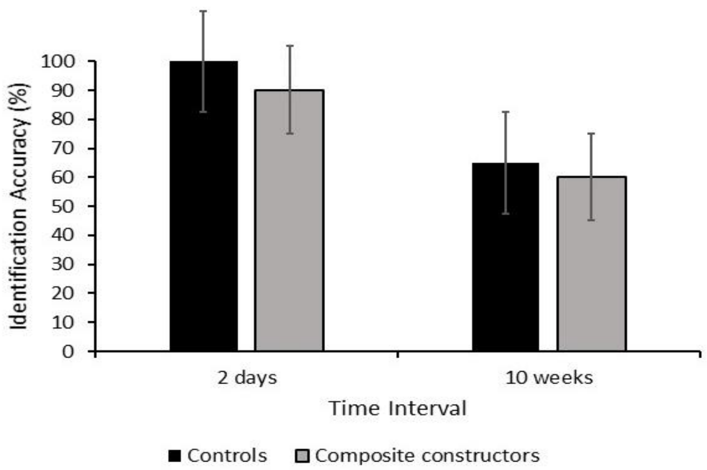

Figure 2.5. Identification accuracy (%) across both composite conditions and time intervals.

Confidence ratings

Overall, confidence ratings were high (*M* = 5.8). A correlational analysis revealed a positive relationship between confidence and recognition accuracy $r(78) = 0.53$, $p < .01$. Confidence was affected by recognition accuracy, whereby participants who made correct identifications were more confident than those who made incorrect identifications $F(1, 78) = 30.29$, $p < .001$, $\eta_p^2 = .28$. There was no effect of composite condition on confidence, $F(1, 76) = .68$, $p = .411$, $\eta_p^2 = .28$. Again, however, a main effect of time delay was found, $F(1, 76) = 31.10$, $p < .001$, $\eta_p^2 = .29$, reflecting lower confidence for identification decisions taken after a ten-week delay. The interaction between these factors was not significant, $F(1, 76) = 2.73$, $p = .10$, $\eta_p^2 = .04$.

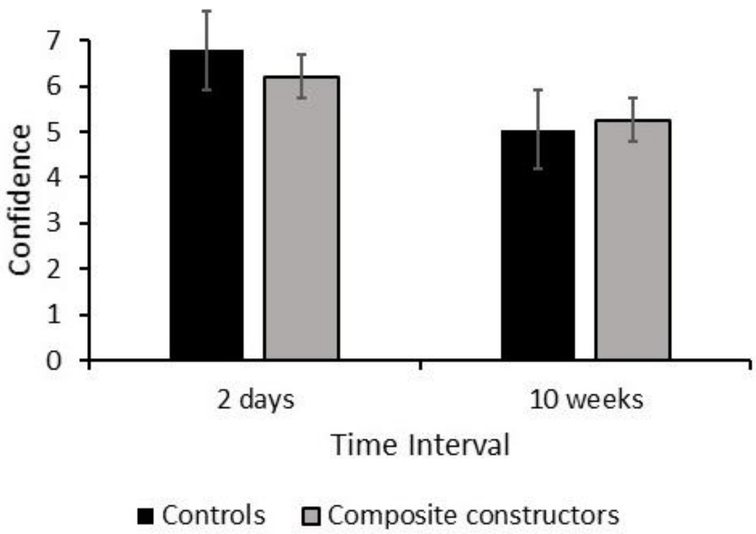

Figure 2.6. Confidence ratings across both composite conditions and time intervals.

Discussion

Previous research has produced conflicting findings regarding the effect of composite construction on subsequent identification performance. Whilst some empirical work has suggested that composite construction incurs a harmful effect, rending participants less able to identify a target (Comish, 1987; Kempen & Tredoux, 2012; Topp-Manriquez et al., 2014;

Wells et al., 2005), other findings document a helpful effect, whereby composite constructors are more able to make subsequent accurate identifications (Davis et al., 2014; Mauldin & Laughery, 1981). Additionally, some research indicates also that composite construction elicits no effect on later recognition accuracy (Davies et al., 1978; Davis et al., 2016; Yu & Geiselman, 1993). Despite these conflicting findings, existing research is unified on the basis that, to date, no studies investigating the effect of composite construction have utilised forensically relevant time delays. The majority of previous research has also employed feature-based composite systems (see Table 2.1). This study therefore sought to extend the investigation of the effect of composite construction, by employing time delays that are more closely aligned with applied settings. Additionally, the current study utilised the more recent holistic composite system EvoFIT. To investigate which of the conflicting documented effects may extend to applied settings, recognition accuracy was compared across composite constructors, and control participants, after either a two-day or ten-week time interval. Participants in the two-day condition viewed a target face and proceeded immediately onto a cognitive interview and composite construction, with recognition accuracy being assessed two days later. In the ten-week condition, participants viewed a target face, then completed a cognitive interview and composite construction 24-48 hours later. Recognition accuracy was subsequently tested ten weeks later.

In this instance, when a two-day delay was employed between target exposure and a subsequent target-present line-up task, recognition accuracy did not differ substantially between composite constructors (90%) and controls (100%). Likewise, when an extended time delay of ten weeks was employed between exposure and recognition, composite constructors (60%) and control participants (65%) were equally able to make accurate identifications from a target-present line-up. This finding is consistent with early research which found that the recognition accuracy of composite constructors does not differ from control participants

(Davies et al., 1978). The finding of a null effect is also in line with additional previous research (e.g., Yu & Geiselman, 1993), including more recent work employing a holistic composite system (Davis et al., 2016).

However, the findings of the current study contrast with previous research which has documented a harmful effect (e.g., Wells et al., 2005). It is possible that this contrast arises as a result of the composite system employed. To date, all studies documenting a harmful effect of construction on recognition have employed feature-based composite systems, which are thought to produce less recognisable images than holistic systems (Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005). Previous research has also found a correlation between composite quality and line-up performance, suggesting that poor-quality composites are more likely to negatively affect identification performance (Wells et al., 2005). A similar finding is also evident in Davis et al.'s (2016) study, whereby facial composites created by children were rated as less recognisable than composites created by adults. In this instance, children, but not adults, experienced a harmful effect of composite construction, and were subsequently less able to accurately identify the target (Davis et al., 2016). Taken together, these findings indicate that poor-quality composites may adversely impact a composite constructors' memory. Considering that holistic systems generate more recognisable images (Frowd, Pitchford, et al., 2012; Solomon et al., 2012), it is possible that the harmful effect was not elicited in the current study because the holistic composite system EvoFIT was employed.

Moreover, in the current work, a substantial decline in recognition performance was found over time. Across both composite conditions, recognition accuracy was at 95% after a two-day delay, but decreased to just 63% after a ten-week delay. This finding is in line with the wider literature, which demonstrates that longer retention intervals negatively affect recognition accuracy (Barkowitz & Brigham, 1982; Deffenbacher, Bornstein, McGorty, & Penrod, 2008; Shapiro & Penrod, 1986; Shepherd & Ellis, 1973). Additionally, participants who made accurate identifications indicated higher confidence levels compared to those who made incorrect identifications. This supports the notion that confidence may act as a useful indicator of accuracy (Palmer, Brewer, Weber, & Nagesh, 2013).

These findings have implications for applied settings. Although previous research has produced conflicting findings regarding the effects of composite construction, the current study suggests that neither the helpful or the harmful effect applies to the holistic composite system EvoFIT. This finding is strengthened by the fact that the recognition accuracy of composite constructors and controls was tested over two distinct conditions, comprising different time intervals. With holistic composite systems increasingly becoming the standard, both in the UK and overseas (Frowd, Pitchford, et al., 2012; Solomon et al., 2012), the findings of the current study are promising, and suggest that, composite construction using such systems is a suitable task for eyewitnesses.

Having discovered that composite construction does not incur any subsequent effects on recognition accuracy, the next chapter investigates a way of improving composite naming rates. Chapter 3 explores the possibility that not all eyewitnesses are equally equipped for the task of composite construction, and aims to provide a means of determining which eyewitnesses are most likely to create good-quality facial composites, based on individual differences in face recognition abilities.

Chapter 3

Individual Differences in Composite Construction

Introduction

The previous chapter investigated the effect of composite construction on subsequent eyewitness identification accuracy. Although the available evidence provides contrasting findings, with both helpful (Davis et al., 2014; Mauldin & Laughery, 1981), and harmful effects (Comish, 1987; Kempen & Tredoux, 2012; Topp-Manriquez et al., 2014; Wells et al., 2005) being documented, Experiment 1 suggests that neither of these effects are elicited when the holistic composite system EvoFIT is employed. Whilst the absence of a harmful effect is a promising finding for applied settings, research nevertheless suggests that composite naming is often poor (see, e.g., Davies et al., 2000; Frowd et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Valentine et al., 2010). The typically low naming rates elicited through studies of composite recognition indicates that facial composites often do not accurately resemble a target face. Although composite quality is thought to have improved in recent years through the production of holistic systems (Davis, Gibson, & Solomon, 2017; Frowd, 2017), one aspect of composite construction that has received limited attention is the role of individual differences.

Facial composites are only necessary in criminal investigations in which the perpetrator is unknown. The process of composite construction therefore arguably relies on unfamiliar face recognition and face memory. The wider face recognition literature provides a wealth of evidence demonstrating that recognition ability for unfamiliar faces varies widely between individuals (Bindemann, Avetisyan, et al., 2012; Bobak et al., 2016; Burton et al., 2010; Russell et al., 2009). The wide range of individual differences is evident on, for example, standardised tests of unfamiliar face recognition, such as the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006). The CFMT requires subjects to identify one of six unfamiliar target faces on test items consisting of a three-face array. Whilst the average score is 80% (Duchaine & Nakayama, 2006), individual scores can range from 51% (Duchaine & Nakayama, 2006) to 100% (Russell et al., 2009).

Although such variation highlights a broad distribution of abilities, an individual's recognition ability is thought to remain relatively stable across different tasks. For example, when presented with photographs of celebrity identities before they were famous, often as children, individuals who perform at ceiling on the CFMT are also able to identify and name more celebrities, outperforming their counterparts (Russell et al., 2009). Likewise, poor performance on the CFMT is indicative of poor familiar face recognition, evidenced by lower naming rates of famous faces (Duchaine et al., 2007). The relative stability of these skills is further supported by research which suggests that face recognition abilities correspond to face matching abilities. On face matching tasks, whereby subjects are required to indicate whether a face pair is the same person (match) or different people (mismatch), individuals with extraordinary face recognition abilities also demonstrate superior skills on tests of both familiar and unfamiliar face matching (Bobak et al., 2016; Robertson et al., 2016). In addition, recognition skills appear to remain relatively stable across time, with face recognition training bearing little impact on accuracy (White et al., 2014; Wilmer, 2017).

Available evidence also suggests that face recognition has a genetic basis (Duchaine et al., 2007; McKone & Palermo, 2010; Wilmer et al., 2010; Zhu et al., 2010). This heritability is evident in the finding that the recognition scores of monozygotic twins are more highly correlated than the scores of dizygotic twins (Wilmer et al., 2010; Zhu et al., 2010). Taken together, these findings therefore converge to demonstrate that face recognition has a genetic basis (Wilmer et al., 2010; Zhu et al., 2010), remains relatively stable across different tasks (Bobak et al., 2016; Duchaine et al., 2007; Robertson et al., 2016; Russell et al., 2009), and across time (White et al., 2014; Wilmer, 2017). Although these individual differences have been studied theoretically, the range and relative stability of face identification also has useful implications for applied, forensic work.

Indeed, previous work has begun to illustrate an association between visual processing and eyewitness identification accuracy. It has been suggested, for example, that individuals with a greater global processing bias are significantly more likely to make correct eyewitness identifications than individuals with a lesser global processing bias (Darling, Martin, Hellmann, & Memond, 2009). Additionally, performance on standardised recognition tests may provide a valuable insight into eyewitness identification accuracy, and help to prevent *mis*identifications, which are currently the leading cause of wrongful convictions (Innocence Project, 2016). A recent study investigated this notion by showing participants a video of a staged crime and subsequently asking them to identify the perpetrator from a ten-face line-up (Bindemann, Brown, et al., 2012). Following the eyewitness test, participants were required to complete a face test, comprised of 40 trials of the 1-in-10 task (Bruce et al., 1999). The 1-in-10 task requires participants to study a single target face and then indicate whether the target is present or absent in a subsequent ten-face photographic line-up. Performance on the two tests were positively correlated, with the percentage of correct identifications on the 1-in-10 task correlating well with the probability of making an accurate eyewitness identification (Bindemann, Brown, et al., 2012). In addition, individuals who excelled on the 1-in-10 task, achieving scores of 90% or more, were always accurate in their eyewitness identification (Bindemann, Brown, et al., 2012).

A similar relationship has been found between performance on a standardised face test and eyewitness accuracy among military personnel (Morgan et al., 2007). In the United States, active duty military personnel are subjected to an interrogation process as part of their training. In this experiential phase of training, recruits are captured, placed in a mock prisoner of war camp (POWC), and subjected to interrogation by an unknown instructor (Morgan et al., 2000, 2007). Recruits experience approximately 30 minutes of exposure to their interrogator, among periods of prolonged isolation, food and sleep deprivation (Morgan et al., 2007). Eyewitness accuracy was investigated under these stressful circumstances by asking recruits to subsequently identify their interrogator from a ten-face sequential line-up. Recruits were also required to complete the Weschler Face Test (1997), which required them to study 24 unfamiliar target faces and subsequently identify each target face from 48 sequentially presented photographs. A positive relationship was found between performance on the two tests, suggesting again, that an individual's ability to perform on a standardised recognition test relates to their aptitude as an eyewitness (Morgan et al., 2007).

The notion that face recognition abilities correspond to eyewitness identification accuracy is further reinforced by a more recent study, which suggests that an individual's performance on the CFMT relates to their eyewitness identification performance (Andersen, Carlson, Carlson, & Gronlund, 2014). In this instance, better recognition abilities, evidenced by higher scores on the CFMT, resulted in more correct identifications in simultaneous, target-present line-ups. Moreover, individuals with higher CFMT scores were also less likely to make a misidentification in target-absent line-ups (Andersen et al., 2014).

The finding that individual differences in facial recognition correspond to eyewitness accuracy (Andersen et al., 2014; Bindemann, Brown, et al., 2012; Morgan et al., 2007) bridges an important gap between theoretical and applied work, by demonstrating that not all eyewitnesses are equally equipped for the task. Considering the utility of standardised recognition tests in these studies, it is possible that these measures may also provide insight into an individual's ability to produce a good-quality facial composite. Although individual differences have not yet been investigated systematically within the composite literature, the finding that composite quality is improved when composites created by multiple eyewitnesses are averaged together (Hasel & Wells, 2007; Valentine et al., 2010), suggests that individuals also vary in their ability to construct accurate composites. This raises the question of whether composite quality can be determined by assessing individual differences in facial recognition. To investigate this question, participants were required to complete two standardised face recognition tests, comprising of the Cambridge Face Memory Test (Duchaine & Nakayama, 2006) and the 1-in-10 task (Bruce et al., 1999). Crucially, both of these tests have been shown to produce large individual differences in recognition performance (Bindemann, Brown, et al., 2012; Duchaine & Nakayama, 2006; Megreya & Burton, 2006; Russell et al., 2009). Participants then proceeded to construct a facial composite of an unknown celebrity using the software system EvoFIT, which is thought to align composite construction more closely with facial recognition (Frowd, 2017). Participant's facial composites were subsequently rank ordered, based on their perceived level of similarity to the target face by a volunteer sample. This experiment therefore provides an indication of whether composite quality can be predicted by an eyewitness' recognition ability which will, in turn, allow police to select only the most suitable eyewitness for the task of composite construction. It is expected that individuals with better recognition abilities will produce more accurate facial composites, comparative to individuals with lower recognition abilities.

Method

Participants

Forty students (7 male, 33 female) from the University of Kent participated in this experiment in return for course credits. The participants ages ranged from 18 to 30 years (*M* = 20.2, *SD* = 2.4). Participants were only eligible for this experiment if they did not watch the soap opera, "Hollyoaks".

Stimuli and procedure

1-in-10 task. In the first stage of this experiment, participants completed forty trials of the 1-in-10 task (Bruce et al., 1999), which has previously been used to assess eyewitness accuracy (Bindemann, Brown, et al., 2012). This task was completed on the software system *PsychoPy* (Peirce, 2007). Each trial presented participants with a single target face, which was 4.1 cm x 5.3 cm in size. Participants were instructed to study the target face until they felt they could identify the target in a subsequent line-up task. The target face was then replaced by a photographic line-up consisting of ten faces, which were 3.5 cm x 4.5 cm in size. Participants were asked to indicate whether the target was present or absent in the line-up and, if present, identify the target by indicating the number corresponding to its position on the keyboard. Each trial consisted of a different target identity. Trials were presented in a randomised order and each participant was given twenty target-present and twenty target-absent trials. In target-present trials, the target identity was presented in the subsequent line-up whereas for target-absent trials, the target identity was omitted. There was no time restriction on any of the trials, each target face and subsequent line-up remained on screen until a keyboard response was given.

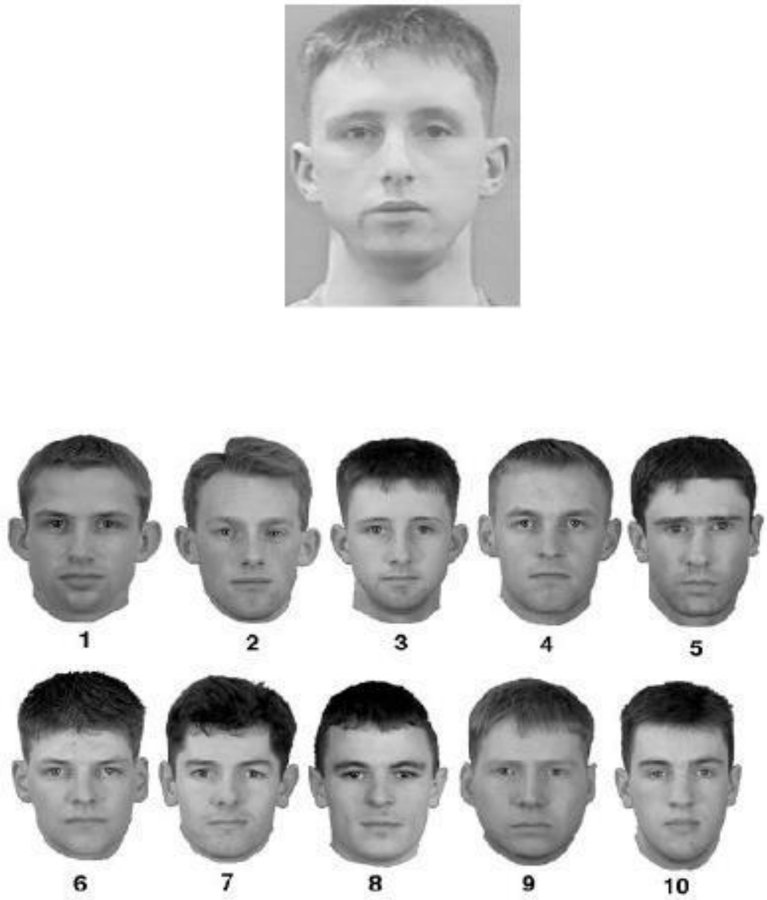

Figure 3.1. Taken from Bruce et al. (1999). An example stimulus of the 1-in-10 task.

Cambridge Face Memory Test. Following the 1-in-10 task, participants completed the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006). The CFMT is a standardised measure of face recognition that is able to effectively assess a broad range of recognition abilities (Duchaine & Nakayama, 2006), and has previously been used to predict eyewitness accuracy (Andersen et al., 2014). The test requires participants to learn six target faces and recognise them on 72 test items which get progressively harder. In the first block of trials, participants are introduced to a target face by viewing a left 1/3 profile, a frontal view, and a right 1/3 profile for three seconds each (see Figure 3.2, row A). Participants are then required to identify the target on a test item, consisting of two distractor faces (see Figure 3.2, row B). Each test item in the first block includes one image identical to that of a study image (e.g. no changes in expression, lighting etc.). In the second block, participants review all six target faces simultaneously for 20 seconds before proceeding to the test items. Again, the test items in this block consist of one target face and two distractors but contain a novel view of the target (see Figure 3.2, row C). The third and final block is similar to block two but Gaussian noise is added to the novel images (see Figure 3.2, row D). On each of the 72 test items, participants indicate their response using the numbers 1-3 on a keyboard. All test items remain on screen until a keyboard response is given.

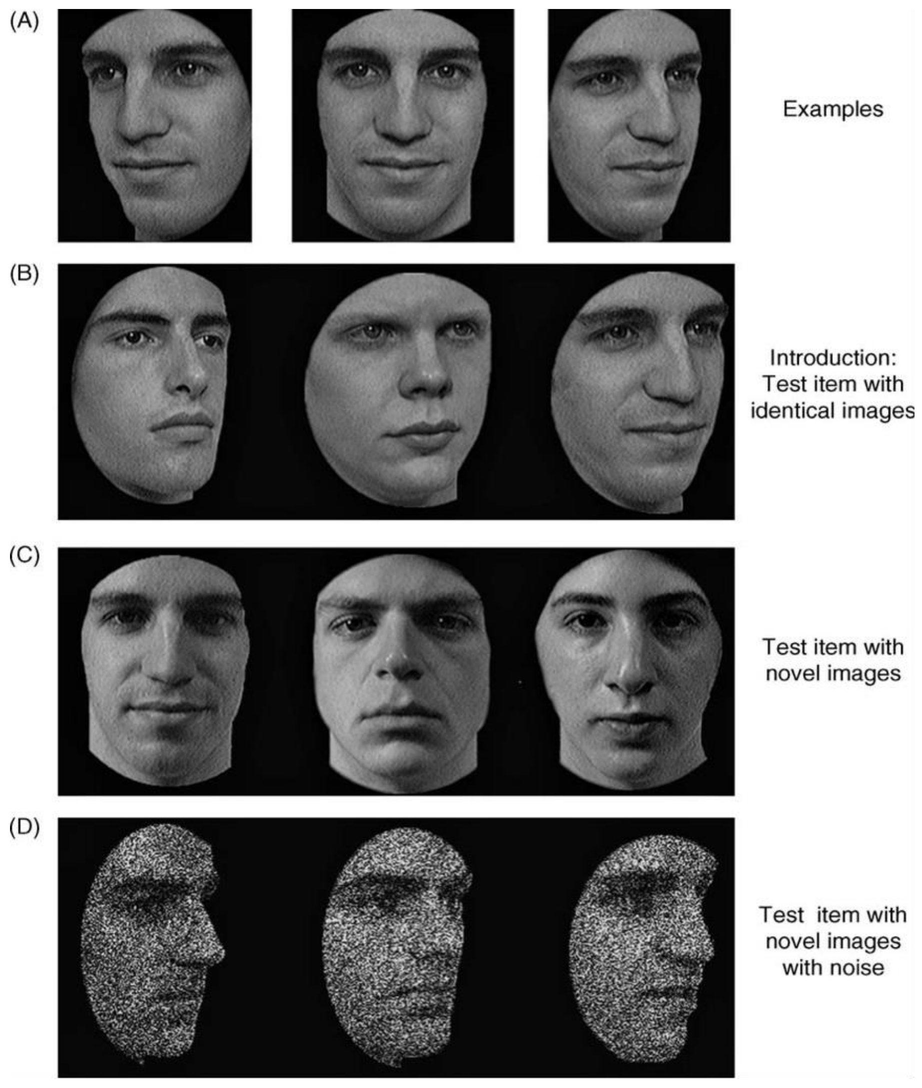

Figure 3.2. Taken from Duchaine and Nakayama (2006). Examples of study and test items in the Cambridge Face Memory Test.

Composite construction. The third and final stage of this experiment required participants to create a facial composite of an unfamiliar celebrity identity. For this purpose, Hollyoaks actors James Sutton (target one) and Greg Wood (target two) were selected as target faces. The eligibility requirements for this study dictated that participants must not watch the soap opera Hollyoaks. This requirement helped to ensure that the target faces would be unknown to the participants, thus allowing construction of an unfamiliar face, as per police procedures with real eyewitnesses. Each participant was shown a photograph of one of the target faces and asked whether they knew who the person was. Target photographs were sourced via a Google search and depicted the target in a full-frontal view, with a neutral expression. Both target photographs were printed in colour and were 7.5 cm x 8.5 cm in size.

Providing the participant was not familiar with the target, they were then required to study the target face for one minute; a timing used elsewhere in the literature for unfamiliar composite construction (e.g. Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Frowd, Carson, Ness, Richardson, et al., 2005). Following this, participants were instructed that they would create a facial composite of the target face and were introduced to the software system, EvoFIT, version 1.6. EvoFIT is a holistic composite system that aligns composite construction more closely with the natural processes involved in face recognition (Frowd, 2017). Composite constructors are presented with a series of whole-face arrays and are required to repeatedly select the faces that they feel best represent the target identity. The process focuses initially on facial shape, whereby constructors make selections based on how accurately each face depicts the size and placement of the target's features. Constructors then progress onto facial texture, and are required to choose the faces that most accurately represent the target, based on differences between the greyscale colouring of the eyebrows, eyes and skin tone. These selections are subsequently 'bred' together to generate a facial composite (for a review see, Frowd, 2017). The process focuses on the internal facial features to help promote subsequent familiar face recognition and naming rates (see, e.g., Ellis, Shepherd, & Davies, 1979). Once the composite has been evolved, composite constructors can make further adjustments using software tools which allow characteristics such as age, weight and trustworthiness to be manipulated to help improve the overall similarity to the target. For this purpose, the researcher acted as an operator for the system, with 20 participants creating a facial composite of target one, and 20 participants creating a facial composite of target two.

Composite similarity rankings

To establish whether individual differences on the 1-in-10 task and the CFMT correlated with the quality of composites produced, 40 participants (9 male, 31 female) were recruited to rank composite accuracy. Participants ages ranged from 18 to 59 years (M = 24.3, SD = 9.0). Each participant was shown all 40 composites, 20 composites for target one and 20 composites for target two. They were then asked to rank order the similarity of each facial composite to the respective target face from 1 to 20. Participants completed this task with the neutral expression, colour photographs used during the construction stage in full view.

The method of this study required two different samples, so distinct labels are provided to avoid confusion. The first group of participants, who were subjected to the CFMT, the 1-in-10 task and composite construction, will here-on-in be referred to as *composite constructors*. The second group of participants, who were recruited to rank composite similarity, will here-on-in be referred to as *composite rankers*.

Results

Individual differences in recognition performance

CFMT accuracy. The average total score on the CFMT was 56.6, which converts to 78.6% and is comparable with normative results elsewhere in the literature (e.g. 80%; Duchaine & Nakayama, 2006). However, broad individual differences were evident across this test, with overall accuracy ranging from 36.1% to 98.6% between composite constructors (SD = 11.8). Accuracy in the first block of this test was at 98.8%, indicating very few mis-takes, but individual scores ranged from 72.2% to 100.0%, revealing early individual differ-ences in recognition accuracy. Average performance subsequently decreased to 76.8% in block two, with accuracy ranging from just 20.0% to 96.7%. Performance deteriorated again in block three to 65.8%, with individual scores ranging from 29.2% to 100.0%.

1-in-10 Task accuracy. On the 1-in-10 task, average identification accuracy was at 61.2% (SD = 14.9). Once again, broad individual differences were evident across this test, with overall accuracy ranging from 27.5% to 95.0% between composite constructors. For tar-get-present trials, the percentage of correct identifications was 64.5%, with individual scores ranging from 20.0% to 100.0%. Likewise, for target-absent trials, a broad distribution of abil-ities was evident. Correct rejections were made 57.9% of the time, but individual accuracy ranged from 10.0% to 100.0%.

Overall accuracy on CFMT and 1-in-10 Task. To explore whether an individual's overall performance on one recognition test was indicative of overall performance on the other, a correlational analysis was performed with the CFMT and 1-in-10 task (see Figure 3.3). This analysis revealed a positive relationship between overall accuracy on the two tests $r(38) = 0.47$, $p < .01$. However, a clear outlier was evident, with the lowest scoring participant achieving a CFMT score which was 3.6 standard deviations below the mean, and a 1-in-10 score which was 1.3 standard deviations below the mean (see Figure 3.3). The correlational analysis was subsequently re-run, excluding the outlier. In this instance, the positive relationship between overall accuracy on the two recognition tests persisted, $r(37) = 0.44$, $p < .01$. The correlation between the two face recognition tests supports the notion that they are testing the same underlying processes.

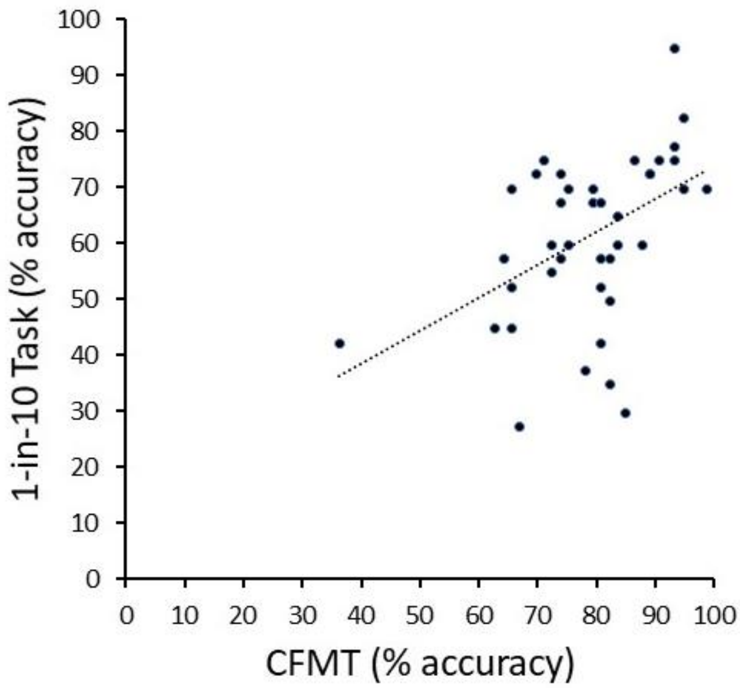

Figure 3.3. Scatter plot for overall performance on the CFMT and 1-in-10 task.

Composite similarity rankings

The quality of a composite constructor's facial composite was determined by calculating the average similarity ranking score across all composite rankers. The facial composites were rank ordered from 1 to 20, therefore lower average scores are indicative of a better-quality composite, which shares a higher level of similarity with the target face. These data are provided in Figure 3.4 and Figure 3.5, and highlight a large variation in composite quality.

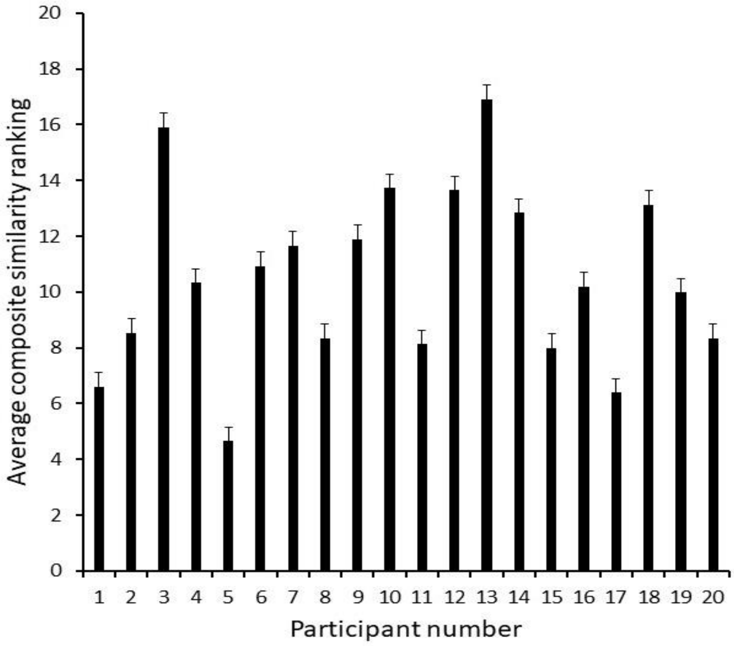

Figure 3.4. *The mean similarity ranking scores for each participant's composite for target one. Error bars represent the standard error of the mean.*

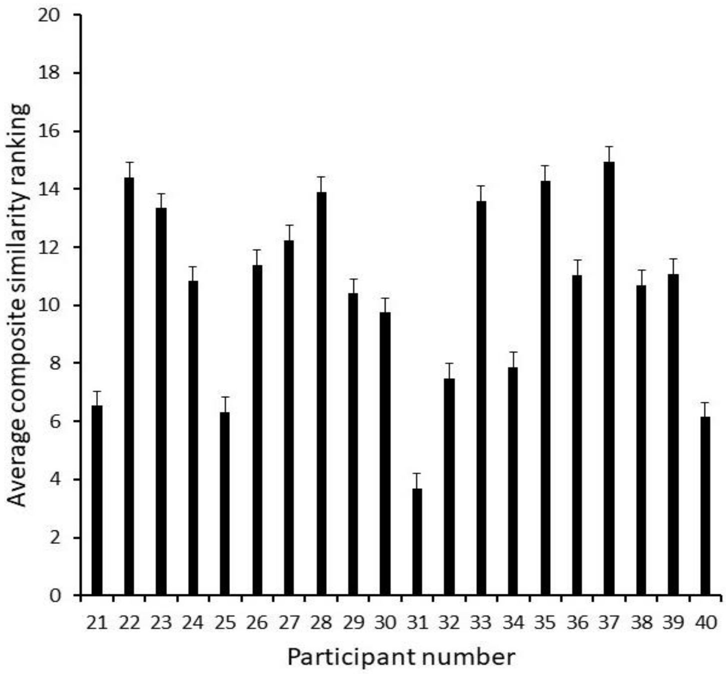

Figure 3.5. *The mean similarity ranking scores for each participant's composite for target two. Error bars represent the standard error of the mean.*

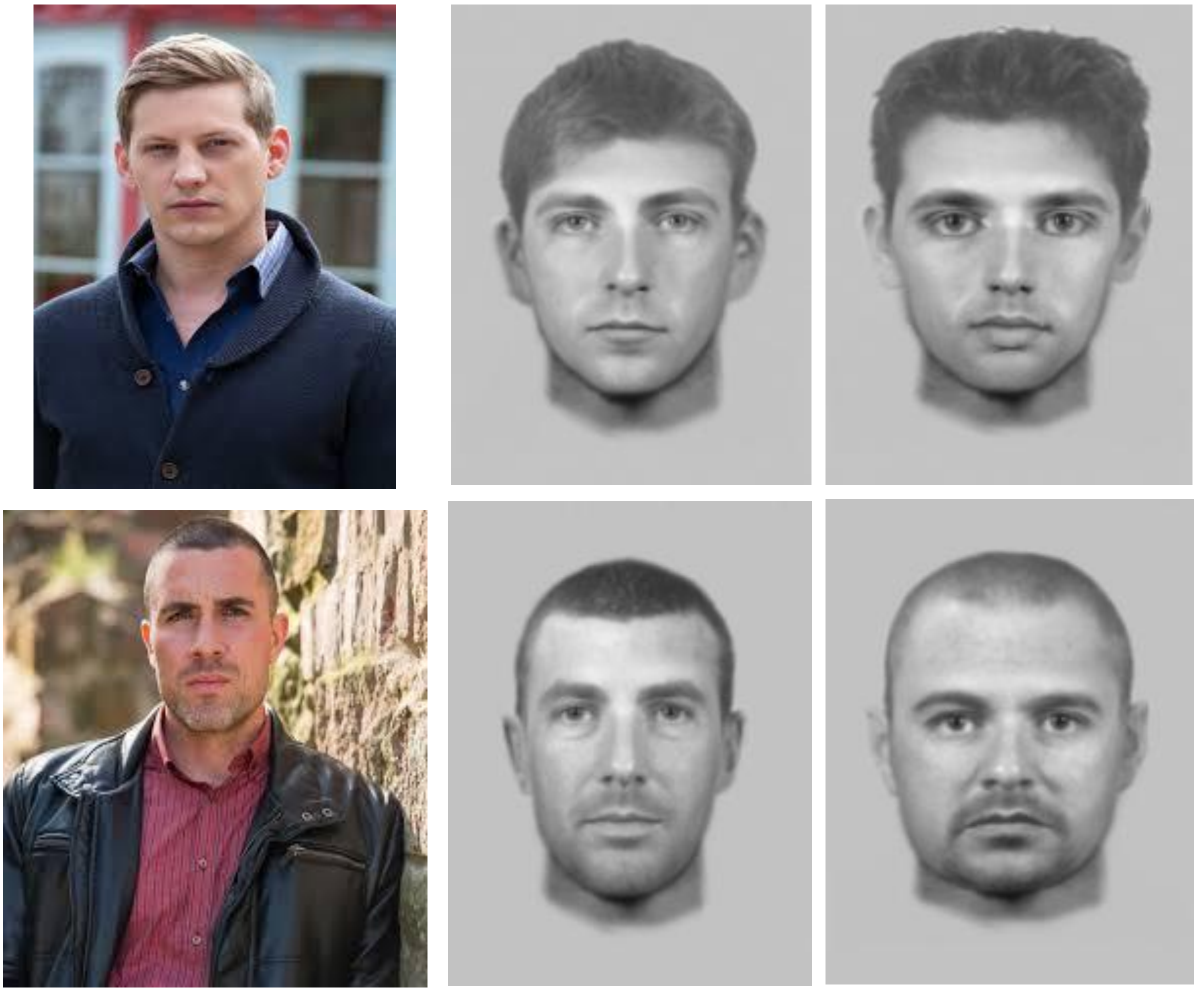

Figure 3.6. Examples of the two target identities (left) and two facial composites created of each identity (center and right). The facial composites represent the best (center) and worst (right) composites, as indicated by their average similarity ratings.

Recognition performance and composite quality

CFMT. To investigate whether individual differences on the CFMT corresponded to composite quality, a correlational analysis was performed (see Figure 3.7). The analysis revealed that an individual's overall performance on the CFMT did not correlate with their average composite similarity ranking score, $r(38) = -.14$, $p = .39$. The data was subsequently broken down for each block of the CFMT and a second correlational analysis was performed, which revealed no significant correlations between composite quality and block one, two and three of the CFMT, $r(38) = -.17$, $p = .31$, $r(38) = -.23$, $p = .16$, and $r(38) = -.02$, $p = .89$, respectively.

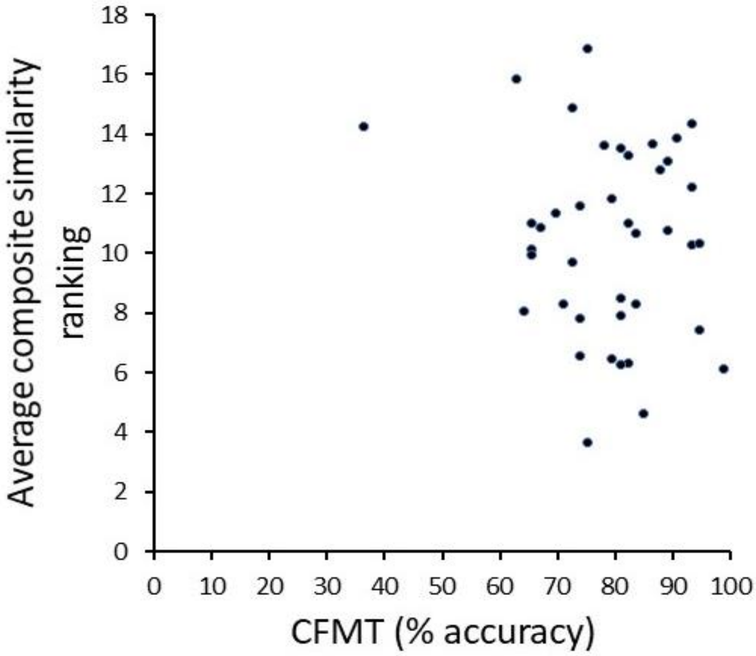

Figure 3.7. A scatter plot for an individual's average composite similarity ranking score and their overall performance on the CFMT.

1-in-10 Task. Likewise, to investigate whether individual differences on the 1-in-10 task corresponded to composite quality, an additional correlational analysis was performed (see Figure 3.8). The analysis revealed that an individual's overall performance on the 1-in-10 task did not correlate with their average composite similarity ranking score, $r(38) = .09$, $p = .58$. The data was then broken down into hits (making a correct identification), misses (incorrectly responding that the target is absent), misidentifications (incorrectly identifying a foil), and correct rejections (correctly responding that the target is absent), to assess whether different aspects of face identification related to composite quality. No significant correlations were found between composite quality and hits, $r(38) =.06$, $p = .70$, misses, $r(38) = -$

.01, p = .94, misidentifications, $r(38) = -.07$, p = .66, or correct rejections, $r(38) = .08$, p = .64.

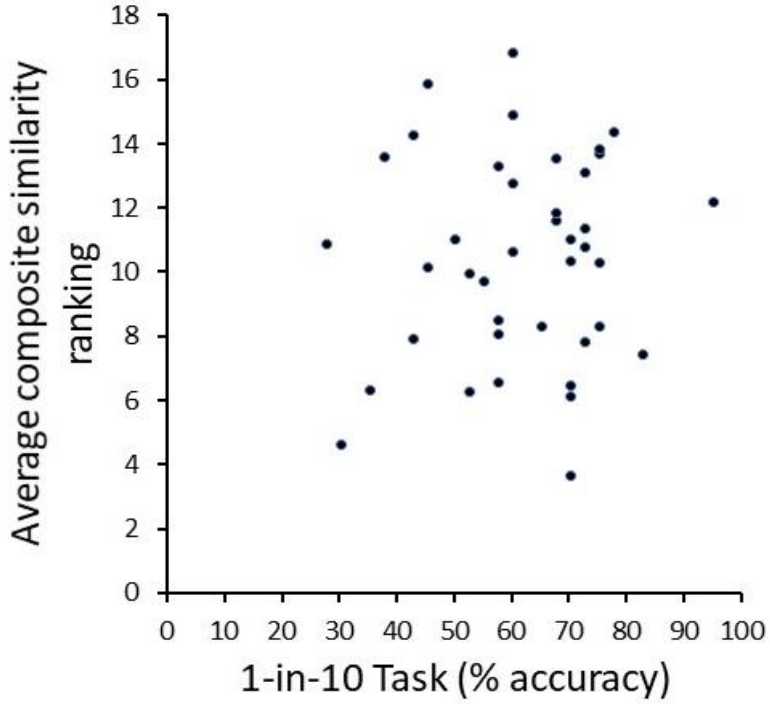

Figure 3.8. A scatter plot for an individual's average composite similarity ranking score and their overall performance on the 1-in-10 task.

General Discussion

Previous research has demonstrated large individual differences in face recognition abilities (Bindemann, Avetisyan, et al., 2012; Bobak et al., 2016; Duchaine & Nakayama, 2006; Duchaine et al., 2007; Robertson et al., 2016; Russell et al., 2009). The broad distribution of these abilities has previously been found to correspond to eyewitness identification accuracy, suggesting that not all eyewitness are equally equipped for the task (Andersen et al., 2014; Bindemann, Brown, et al., 2012; Morgan et al., 2007). This study sought to extend the investigation of individual differences to examine whether the recognition skills of an eyewitness can predict the quality of the facial composite they are able to construct. To investigate this, participants were subjected to two standardised face recognition tests, comprising the CFMT (Duchaine & Nakayama, 2006) and the 1-in-10 task (Bruce et al., 1999), both of which have previously been used to assess eyewitness accuracy (Andersen et al., 2014; Bindemann, Brown, et al., 2012). Participants then proceeded to construct a facial composite of an unknown celebrity identity using the software system EvoFIT. All facial composites were subsequently ranked by a volunteer sample to establish the level of similarity shared between each participant's facial composite and the respective target face.

A moderate correlation was found between an individual's overall performance on the CFMT and their overall performance on the 1-in-10 task, which is comparable in size to other correlations across different tasks of face recognition (Fysh, 2018; Fysh & Bindemann, 2018; McCaffery, Robertson, Young, & Burton, 2018; Robertson, Jenkins, & Burton, 2017). Despite this, an individual's performance on the CFMT and the 1-in-10 task did not correlate with their facial composite's average similarity ranking score. The absence of a correlation in this instance, suggests that composite construction may engage separate processes, that are not currently accounted for in unfamiliar face recognition tasks.

The results of the current study converge with other recent research, which suggests that individual performance on tasks of face memory and face matching are unrelated to an individual's ability to detect a target identity in a crowded scene from a composite image (Bate et al., 2018). Although Bate et al.'s (2018) study focuses on an individual's ability to match a composite image to a target identity, rather than composite quality, these studies are unified on the basis that face recognition abilities and performance on a facial composite task were unrelated in both instances. Given that facial composites are not "real" faces, it is possible that facial composite tasks may not align with other aspects of face recognition and may exploit different cognitive processes.

The CFMT and the 1-in-10 task were selected for the current study because they have both previously been used to assess eyewitness accuracy (Andersen et al., 2014; Bindemann, Brown, et al., 2012). However, both of these standardised tests are recognition tests, requiring the selection of a target identity from a range of different identities. In contrast, the process of composite construction requires eyewitnesses to differentiate between highly similar faces, and select the best likeness to the target face, based on minute differences. The process involved may therefore be more reflective of a *perception* task, rather than a *recognition* task. Face recognition is known to be affected by variability, with the learning of new faces being facilitated by variability (Kramer, Jenkins, Young, & Burton, 2017; Ritchie & Burton, 2017). It is also known that within-person variability can often be large, with photographs of the same identity being incorrectly perceived as different people by individuals who are unfamiliar with the target (Jenkins, White, Van Montfort, & Burton, 2011). However, the process of composite construction exposes eyewitnesses to very little variability *between* identities, as opposed to natural variability *within* an identity, therefore it is possible that composite construction does not accurately align with the process of natural face recognition.

Moreover, in the current work, composite quality was determined by average similarity ranking scores. These similarity rankings were obtained across a group of 40 composite rankers. It has previously been suggested that what constitutes a good likeness to a target face is based on an individual's level of familiarity and representation of the target's identity (Ritchie, Kramer, & Burton, 2018). The composite rankers in this study viewed one photograph of the target identity and subsequently ranked composite quality based on each composite's similarity to the photograph. The composite rankers lack of exposure to variability within the target's identity may therefore have resulted in a lower tolerance for different representations of the target. In addition, because likeness is based on familiarity, it is a rather individual concept, as opposed to a generalised concept that is shared between individuals

(Ritchie et al., 2018). This therefore suggests that the similarity ranking scores obtained here yield a noisy stimulus, that may not be comparable across individuals.

These findings have implications for applied settings. Although previous research has indicated that an individual's face recognition ability relates to their aptitude as an eyewitness (Andersen et al., 2014; Bindemann, Brown, et al., 2012; Morgan et al., 2007), the current study suggests that recognition abilities do not correspond to composite quality. The finding that an individual's recognition ability is unrelated to their ability to construct a good-quality facial composite suggests that composite construction is not a recognition task. This finding is strengthened by the fact that an individual's overall performance on the CFMT and 1-in-10 task did correlate. In criminal investigations with multiple eyewitnesses, it therefore remains unclear which eyewitnesses would be most suited for the task of composite construction. It is possible that real-world face processing tasks, such as composite construction, require a more specific process that is currently undetected by standardised face recognition tasks. Future research should focus on the possibility that the task of composite construction may be more aligned with face perception, as opposed to face recognition. This will allow for more exploration into the processes involved, and may provide a means of improving composite quality through the selection of the most suited eyewitnesses, likely to construct the best quality composites.

In the next chapter, another means of improving the identification of facial composites is investigated. Experiment 3 investigates whether composite naming can be improved through the provision of contextual information and aims to conceptualise *how much* information is required to facilitate the recognition of facial composites.

Chapter 4

Context-Based Recognition of Facial Composites

Introduction

The previous chapter investigated individual differences in composite construction to establish whether some eyewitnesses are better equipped for the task than others, based on differences in face recognition abilities. The absence of a correlation between an individual's recognition ability and composite quality means that it remains unclear which eyewitnesses are most suited for the task of composite construction. Experiment 3 investigates another means of improving composite identification rates by examining the role of contextual information in composite identification.

In criminal investigations, when a facial composite is released to the public, the composite image is typically displayed with context (for examples, see Crimestoppers, 2018; Metropolitan Police, 2016). The context provided may be explicit, detailing the perpetrator's appearance (e.g., age, height, clothing), or implicit, with the composite image being displayed in locations which the offender is known to frequent.

Whilst the process of composite construction is applied widely in police settings, a coherent body of psychological research suggests that it is a difficult task. As a result of this difficulty, facial composites often provide limited resemblance to a target face. This is evident through studies of composite naming, which often present identification rates below 20% (Davies et al., 2000; Frowd et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Valentine et al., 2010). However, akin to the way in which composites are displayed with context in applied settings, many psychological studies inadvertently provide contextual information that may facilitate naming rates. For example, these studies may use paradigms that restrict context by informing participants that the target is famous (e.g., Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Frowd, Carson, Ness, Richardson, et al., 2005), or by revealing additional information about the target, such as their occupation (e.g., Frowd, Pitchford, et al., 2011; Frowd, Skelton, et al., 2011; Valentine et al., 2010). Restricting context in this manner reduces the range of possible identities that a facial composite may represent, which could result in higher naming rates than spontaneous naming that is based on recognition of the facial composite only (see, e.g., Valentine et al., 2010). To date, however, this possibility has not been examined directly.

In contrast, other domains already provide some evidence that context facilitates recognition. The phenomenon of semantic priming, for example, whereby response times are faster if a target is preceded by semantically related primes, is known to affect the speed of word recognition (Collins & Loftus, 1975; Seidenberg et al., 1984). These semantic priming effects extend to familiar face recognition, whereby a person is recognised more quickly following the provision of semantic information (Bruce et al., 1993; Burton et al., 1990). This information can be quite generic, such as a person's nationality or occupation, but will facilitate face recognition nonetheless (Klatzky, Martin, & Kane, 1982). This suggests that even relatively unconstrained context could incur advantages in composite identification.

Although semantic priming effects are typically observed in the recognition of familiar faces, contextual information has also been found to aid the recognition of unfamiliar faces. Recent research, for example, has shown that presenting unfamiliar faces with a single behavioural description can aid subsequent face recognition (Mattarozzi, Colonnello, Russo, & Todorov, 2018). In this instance, presenting faces with either a positive or a negative behavioural description (e.g., "He volunteered to stay late to help a co-worker" or "She insulted a stranger"; Mattarozzi et al., 2018, p.2), resulted in higher recognition accuracy than faces that were presented alone, or with a neutral description. The finding that a single description can facilitate face memory supports previous work investigating the effect of contextual elaboration.

Kerr and Winograd (1982), for example, investigated the effect of contextual elaboration on person recognition by exposing participants to unfamiliar faces presented with none, or one, two or three descriptive phrases. These phrases provided information relating to the target's personality, occupation and hobbies. The provision of a single descriptive phrase was already sufficient context to aid face identification, increasing recognition accuracy from 69% to 79%. However, this effect was not enhanced through the addition of further phrases (recognition accuracy of 76% and 75% in the two and three phrase conditions, respectively). This points to an all-or-none response to context, as opposed to a graded increase, for facilitating face recognition relative to the amount of context provided. To explore the effect of contextual elaboration further, Kerr and Winograd (1982) examined the role of context *reinstatement,* by exposing participants to the same context both during initial face encoding and subsequent recognition. This revealed that contextual reinstatement can augment recognition further (Kerr & Winograd, 1982).

Context reinstatement has also been found to facilitate recognition in field experiments investigating eyewitness accuracy. Early research in this domain asked shop assistants to identify a previously encountered customer from either two hours or 24 hours earlier (Krafka & Penrod, 1985). A physical and mental cue provided context reinstatement, by presenting assistants with a non-photographic form of the customer's identification, such as a signed cheque, and asking assistants to recall the previous interaction with the customer. In this instance, the ability to accurately identify the customer from a photographic line-up increased from 27% to 60% when contextual information was reinstated after a two hour delay, and from 30% to 50% after a 24 hour delay (Krafka & Penrod, 1985).

Overall, these findings converge to suggest that context facilitates face recognition (Bruce et al., 1993; Burton et al., 1990; Kerr & Winograd, 1982; Kraflca & Penrod, 1985; Mattarozzi et al., 2018), particularly through the reinstatement of a previous context at the time of recognition (Kerr & Winograd, 1982; Krafka & Penrod, 1985). Considering that facial composites are typically displayed with context in applied settings, and research settings suggest that spontaneous composite naming is difficult (Frowd et al., 2014; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Valentine et al., 2010), this raises the question of whether similar context effects exist in composite recognition. To investigate this question, participants were asked to name facial composites of celebrity identities in the current study. The amount of contextual information that was provided with these facial composites was varied systematically between conditions and consisted of either no additional context, a "minimal" context condition comprising a single, generic piece of semantic information about a person (such as occupation), or a "maximum" context condition, in which a set of semantic information was provided that narrowed down the number of possible target identities substantially (e.g., a combination of nationality, occupation, age and filmography). This experiment therefore provides an indication of whether composite naming rates can be improved by providing contextual information, and also aims to provide a measure of *how much* context is required to facilitate composite recognition. In addition, a context-only condition was utilised, in which no facial composite was provided, to determine how effective the context is for person identification on its own.

Method

Participants

Eighty students (63 female, 17 male) from the University of Kent, with a mean age of 20 years (*SD* = 4.8; range 18-55), participated in this experiment in return for course credit. Four groups of observers were tested on a between-subjects basis (each group with $N = 20$).

Stimuli

Facial composites. Eight actors (Benedict Cumberbatch, Zac Efron, Ryan Gosling, Tom Hardy, Tom Hiddleston, Theo James, Daniel Radcliffe and Channing Tatum) and eight international footballers (Steven Gerrard, Joe Hart, Jordan Henderson, Adam Lallana, Frank Lampard, Wayne Rooney, John Terry and Jamie Vardy) were selected as target faces. At the time of choosing, the target identities had an average age of 33.6 years (*SD* = 4.6). A full-frontal, neutral expression colour photograph of each target identity was sourced via an internet search, using the celebrity's name as the search term. The first author then constructed a facial composite of each target face with the respective target photograph in full view.

All facial composites were constructed on EvoFIT using the software version 1.6. EvoFIT is a holistic composite system that has been the focus of extensive research and aims to produce more identifiable images that elicit higher correct naming (e.g., Frowd, 2012; Frowd et al., 2012; Frowd, 2017). The process requires composite constructors to view a number of screens, each depicting 18 whole faces, and to repeatedly select the faces that most accurately represent the target identity. Initially, selections are based on facial shape, with the composite constructor selecting the images that most accurately resemble the shape and location of the target's features. Subsequent selections are based on facial texture, with constructors selecting images based on differences between the greyscale colouring of eyebrows, eyes and skin tone (Frowd et al., 2014). Using principle components analysis (PCA) and an evolutionary algorithm, these selections are subsequently 'bred' together to evolve into a facial composite. Finally, software tools are used to adjust characteristics such as age, weight and masculinity to improve the overall likeness of the composite image to the target face (Frowd, 2017; Frowd et al., 2004).

Procedure

The experiment was conducted on a desktop computer equipped with *PsychoPy* software (Peirce, 2007) and employed four different context conditions, comprising no-context, minimal context, maximum context and context-only. In the no-context condition, minimal context and maximum context conditions, participants were informed that they would be viewing computer generated images of different individuals, and were instructed to try and name each image. Participants were informed that, if they could not name the person, they should type "unknown" in order to move onto the next target identity. The facial composites were presented in a randomised order, which varied for each participant, and between conditions. The amount of information that a facial composite was displayed with also varied between conditions. In the no-context condition, each facial composite was simply displayed alongside the question "Can you identify and name this person?". The minimal context condition provided information relating to the target's occupation, by displaying the composite either alongside the question "Can you identify and name this footballer?" or "Can you identify and name this actor?". The maximum context condition provided three to four pieces of semantic information alongside each composite relating to the target's identity, such as their age, occupation, football-playing position and club, or filmography (for examples, see Figure 4.1). In the context-only condition, participants were provided with the same semantic information as in the maximum context condition, but did *not* view a facial composite.

In accordance with the wider composite naming literature (e.g., Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Frowd, Carson, Ness, Richardson, et al., 2005), the 'tip-of-the-tongue' phenomenon was acknowledged. Thereby, if participants were unable to recall a name, an unambiguous description of a target was accepted. In the maximum context condition, participants had to provide additional information that was not displayed alongside the composite for such a description to be accepted.

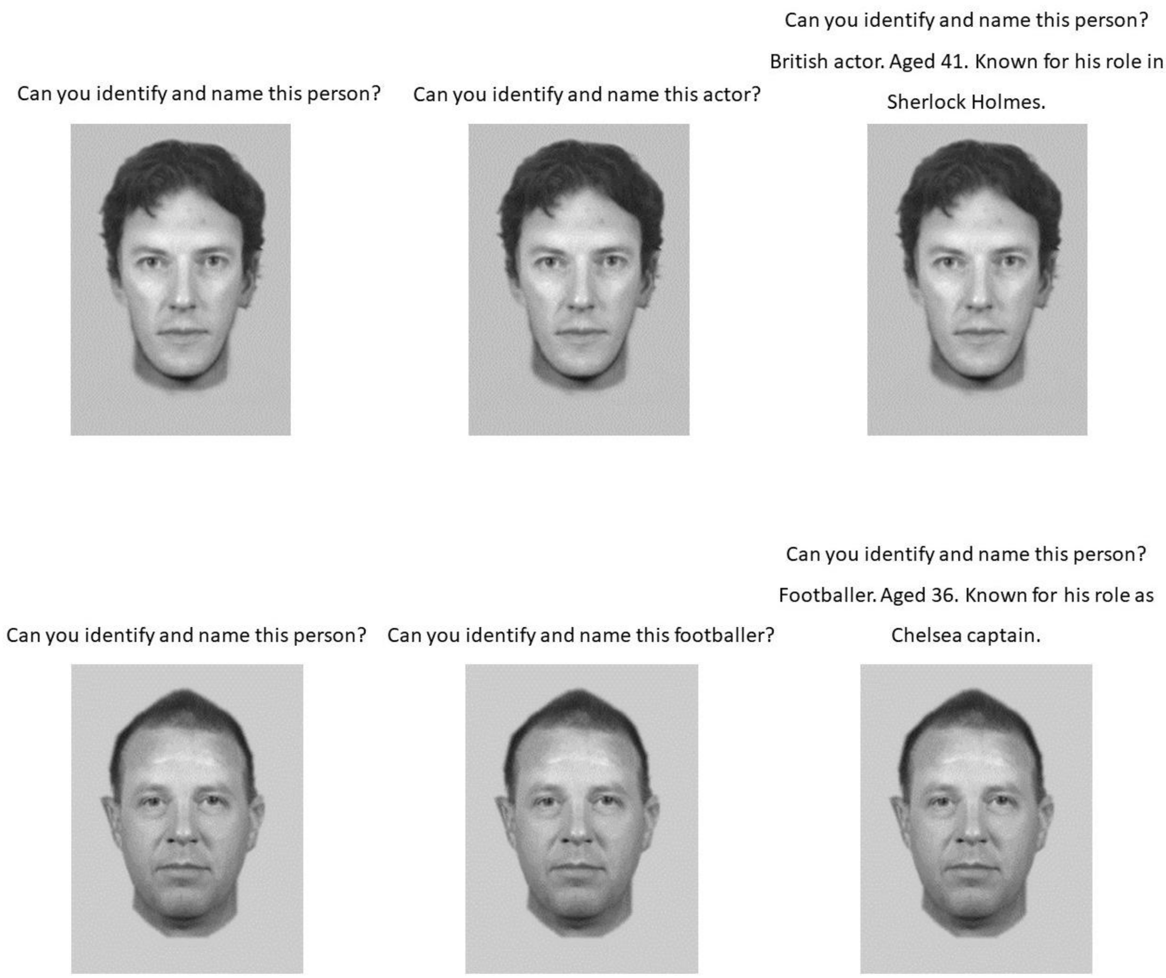

Figure 4.1. Examples of the actor, Benedict Cumberbatch (top), and footballer, John Terry (bottom), in the no context, minimal context and maximum context conditions (left to right).

This experiment employed a staggered design, in which four groups of observers were tested on a between-subject basis (each group with N = 20). Group 1 completed three context conditions, comprising no-context, minimal context, and maximum context, which were always applied in this order. This allowed for assessment on a within-subject basis whether the facial composites that observers cannot name without context can be identified when minimal or maximum context is provided. Group 2 completed a variant of this manipulation, by completing the minimal followed by the maximum context conditions. Group 3 only completed the maximum context condition, whilst Group 4 completed the context-only condition, followed by the maximum-context condition (see Table 4.1). Groups 1, 2 and 4, who experienced multiple context conditions were unaware that the same stimuli would be seen in each condition. In addition to the within-subject comparison of the context conditions, this design therefore also allowed for a direct comparison of accuracy across the four context conditions on a between-subject basis, based on the first condition that each group was given, that is uncontaminated by any previous exposure to the composites.

Table 4.1. The assigned context conditions for each group.

	Context Only	No Context	Minimal Context	Maximum Context
Group 1		✓	✓	✓
Group 2			✓	✓
Group 3				✓
Group 4	✓			✓

On completion of their assigned context conditions, all participants were provided with a familiarity check to establish how many of the target identities were actually known to them. This consisted of the 16 famous-face photographs that had been employed for composite construction, which participants were asked to identify. The target photographs were presented in a randomised order, which varied between participants.

Results

Familiarity check

Participants accuracy was assessed only for targets that could be identified in the familiarity check. Thus, conditional naming rates were calculated, per participant, by dividing the number of correctly named composites in each context condition by the number of correctly named photographs of the same people in the familiarity check. This procedure is a standard measure used elsewhere in the literature (e.g., Frowd, Carson, Ness, Richardson, et al., 2005; Frowd et al., 2014). Two participants were not able to identify any target identities in the familiarity check and therefore had their data discarded and replaced (*N* = 2), based on a procedure adopted elsewhere in the literature (e.g., Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005). For the remaining participants, substantial variation in target familiarity was evident, with between 6% and 85% of target photographs named correctly across observers (M = 39.1%, SD = 26.3%).

Between-subjects comparison of context

Naming rates (% accuracy) were then compared across the context conditions. These data are provided in Table 4.2 and suggest clear context effects. For example, these data show that accuracy was at 5.4% with no context, at 8.6% to 11.8% in the minimal context condition for Group 1 and 2, and in excess of 55% with maximum context across all groups.

Table 4.2. Naming rate percentages across all context conditions and groups.

	Context Only	No Context	Minimal Context	Maximum Context
Group 1		5.4	11.8	56.0
Group 2			8.6	56.6
Group 3				55.6
Group 4	65.5			61.8

To analyse these data, a one-way between-subjects ANOVA was conducted, based on the data of the first context condition that was shown to each of the four groups of participants (no-context vs. minimal context vs. maximum context vs. context-only). This revealed a significant effect of context condition on composite naming rates, $F(3, 76) = 45.11$, $p <$ .001 $\eta_p^2 = .64$. A series of independent samples *t*-tests (with alpha corrected to .05/6 = .008

for six comparisons) revealed that composite naming rates were higher in the maximum context condition compared to the no-context and minimal context conditions, $t(38) = 6.90$, $p < .001$, $d = 2.14$ and $t(38) = 6.34$, $p < .001$, $d = 1.97$, respectively. By contrast, composite naming rates were comparable for the no-context and minimal context conditions, $t(38) = 0.98$, $p = .335$, $d = 0.30$. This pattern of results demonstrates that contextual information facilitated composite naming, particularly when more than minimal context is provided.

The results also show that composite naming rates were higher in the context-only condition than in the no-context condition, $t(38) = 10.70$, $p < .001$, $d = 3.32$, and the minimal context condition, $t(38) = 9.85$, $p < .001$, $d = 3.05$. Moreover, accuracy was comparable for the context-only and maximum context conditions, $t(38) = 1.15$, $p = .260$, $d = 0.35$. Thus, the provision of facial composites did not appear to boost naming rates additionally under maximum context, providing further evidence for the powerful influence of this manipulation.

Within-subject comparison of context

To confirm the pattern of the between-subjects analysis, the data was also analysed separately for participant Group 1, 2 and 4, who were exposed to multiple context conditions (see Table 4.1). For Group 1, which was exposed to the no-context, minimal context and maximum context conditions, ANOVA revealed an effect of context condition on composite naming rates, $F(2, 38) = 39.89$, $p < .001$, $\eta_p^2 = .68$. Paired-sample *t*-tests (with alpha corrected to .05/2 = .025 for two comparisons) revealed that composite naming rates were higher in the minimal context than the no-context condition, $t(19) = 2.49$, $p < .025$, $d = 0.56$, and the maximum context than the minimal context condition, $t(19) = 5.93$, $p < .001$, $d = 1.33$. Similarly, for Group 2, which was exposed to the minimal and maximum context conditions, naming rates improved when context was increased, $t(19) = 9.83$, $p < .001$, $d = 2.20$. In contrast, composite naming rates were comparable for the context-only and maximum context conditions for Group 4, $t(19) = 0.84$, $p = .41$, $d = 0.19$. Considered together, these data confirm the pattern of the between-subject analysis by demonstrating that provision of context substantially increases composite naming rates. In the current study, this effect is such that person recognition is not supported further by the presentation of a facial composite when maximum context is provided.

General Discussion

Whilst existing literature suggests that context can aid recognition of both familiar (Bruce et al., 1993; Burton et al., 1990) and unfamiliar faces (Kerr & Winograd, 1982; Mattarozzi et al., 2018), the effect of context on the naming of criminal facial composites has not been investigated systematically. This study therefore examined whether the facilitating effect of context on recognition extends to composite naming. To investigate this, participants were required to name facial composites constructed of celebrity identities. These composites were accompanied by varying levels of context, which progressed from providing no additional information (no-context), to revealing the target's occupation (minimal context), and then to providing several items of semantic information, such as age, occupation, and what the target was known for (maximum context).

Composite naming when no contextual information was provided was very low, at 5% accuracy, akin to spontaneous naming in other studies (see, e.g., Frowd et al., 2014; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Valentine et al., 2010). However, in line with the wider literature also, the provision of context facilitated recognition substantially (e.g., Bruce et al., 1993; Burton et al., 1990; Kerr & Winograd, 1982; Kraflca & Penrod, 1985). For example, composite naming rates improved to approximately 10% with minimal context, and to over 55% when maximum context was given. This increase in the maximum context condition in particular indicates that the provision of three items of semantic information about a target's identity can dramatically improve the recognition of facial composites. However, the increase from 5% to approximately 10% between the no-context and minimal context conditions is relatively low, indicating that previous research providing limited additional information about a target, such as occupation, is unlikely to have greatly inflated naming rates (e.g., Frowd, Pitchford, et al., 2011; Frowd, Skelton, et al., 2011; Valentine et al., 2010).

These substantial differences between the minimal and maximum context conditions provide an interesting finding in light of Kerr and Winograd's (1982) study, where the context of one descriptive piece of semantic information about a person facilitated recognition, but performance was not enhanced with the addition of further phrases. It is possible that this contrast in findings is elicited by the difference in stimuli, as Kerr and Winograd's (1982) participants viewed facial photographs rather than facial composites. This suggests that facial composites may require more contextual information in order to be recognised, compared to photographs. In turn, considering that composites often provide more limited similarity to a target, and are consequently more difficult to name compared to photographs (see e.g., Frowd et al., 2014), this would suggest also that context is particularly beneficial when facial information is compromised.

The current experiment also provides evidence to support this reasoning, as participants were able to identify a target equally well when provided with maximum context, regardless of whether a facial composite was presented or not. This suggests that three or four items of person-related semantic information, of the type that was employed under maximum context here, can provide a more effective means for person identification than a facial composite. It is also possible, however, that the absence of a difference in recognition performance between the maximum context and context-only conditions arises from the quality of facial composites that were employed here. This notion is arguably supported by the mean target naming within the current study. In this instance, correct naming rates when minimal context was provided were at approximately 10%, which is lower than some comparable studies. For example, recent work employing EvoFIT has achieved mean correct naming rates of 45% when participants were informed that they would be naming facial composites of international footballers (Frowd et al., 2012). Providing this level of context is comparable to the minimal context condition employed within the current study, therefore suggesting that differences in methodology may have negatively affected composite quality in this instance.

Most notably, within the current study, facial composites were created with the target photograph in full view. This method of composite construction differs substantially from the majority of contemporary composite naming research, which employs a delay between target exposure and composite construction. Indeed, in Frowd et al.'s (2012) study, a retention interval of 24 hours was utilised, which is also more closely aligned with the delay implemented in applied settings (Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005). Although it might be expected that constructing a facial composite with the target in view would produce a better-quality composite than those constructed from memory, it is possible that this method of construction resulted in featural processing whereby considerable emphasis was placed on individual facial features. In turn, this feature-based approach to construction may have interfered with EvoFIT's holistic approach to construction, therefore resulting in poorer quality composites. Thus, it is currently unclear whether the recognition of composites that provide a very good likeness to a target would still be enhanced with the type of maximum context that was given here. To address this issue, future research should examine whether similar contextual effects extend to composites that have been created from memory.

Nevertheless, these findings have implications for applied settings. The use of facial composites in criminal investigations is widespread (see, e.g., Frowd, Pitchford, et al., 2012;

Solomon et al., 2012), but composites are not always named correctly (Davies et al., 2000; Frowd et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Valentine et al., 2010). The current study highlights a potential means of improving composite naming dramatically, by demonstrating the powerful facilitating influence of context. Two potential implications arise from this. Firstly, if facial composites are recognised through the provision of contextual information, as opposed to the composite image *per se*, then this may provide less compelling evidence. Jurors, for example, may be less inclined to believe an eyewitness' account if it was evident that the *information* rather than the *person* had been recognised. Secondly, the maximum context and context-only conditions elicited the highest naming rates within this study, therefore suggesting that displaying a facial composite with at least three pieces of contextual information would be the most beneficial procedure for criminal investigations. It is recognised that this recommendation is based on the type of context that can be provided for famous faces (e.g., occupation, nationality, filmography). Future research must explore the type of contextual information that is suited best to eliciting recognition of unfamiliar faces.

Chapter 5

Summary, Conclusions and Future Research

5.1 Summary and conclusions

This thesis investigated the use of facial composites in person identification. The first chapter began by providing an overview of how facial composites are utilised in applied settings and outlined the development of composite systems across four generations. Research has consistently shown that feature-based systems are capable of producing correct naming rates with a mean of 20% when short retention intervals are employed between target exposure and composite construction (e.g. Bruce et al., 2002; Davies et al., 2000). However, when forensically relevant delays are employed, naming rates decrease to just 3% (Frowd et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005). The development of holistic composite systems has since aided identification rates substantially, and these systems are now becoming the standard, both in the UK and overseas (Frowd, 2012; Frowd et al., 2012; Solomon et al., 2012).

Considering the increasingly wide-scale usage of facial composites, the question of whether composite construction is a suitable procedure for real eyewitnesses arises. The notion that composite construction elicits subsequent effects on an eyewitness' identification performance receives support from a variety of studies. It has been suggested, for example, that the process of composite construction incurs a *harmful* effect, rendering an eyewitness subsequently less able to identify the original target (Comish, 1987; Kempen & Tredoux, 2012; Topp-Manriquez et al., 2014; Wells et al., 2005). In stark contrast, a *helpful* effect has also been found, whereby eyewitnesses are more likely to make accurate identifications following composite construction (Davis et al., 2014; Mauldin & Laughery, 1981). Moreover, some studies have also indicated that composite construction elicits *no* subsequent effect on identification performance (Davies et al., 1978; Davis et al., 2016; Yu & Geiselman, 1993).

Together, these findings demonstrate complex and conflicting results, meaning that it currently remains unclear whether real eyewitnesses should participate in the process of composite construction.

Chapter 2 investigated the effect of composite construction on subsequent line-up performance, and aimed to test this over forensically relevant delays, which are more closely aligned with applied settings. Existing research in this domain has often used very short time delays between exposure to a target identity and composite construction, with many studies employing no delay between the two tasks (e.g., Comish, 1987; Davis et al., 2014, 2016; Yu & Geiselman, 1993). Likewise, relatively short time delays have also been employed between composite construction and subsequent recognition tasks, with intervals ranging from no delay (e.g., Davies et al., 1978; Davis et al., 2014; Mauldin & Laughery, 1981), to a maximum delay of one week (e.g., Topp-Manriquez et al., 2014). Experiment 1 therefore aimed to examine the effect of composite construction over more meaningful time delays that reflect the intervals an eyewitness may experience in a real-life criminal investigation. For this purpose, a delay of either two days or ten weeks was employed between exposure to a target face and a subsequent line-up task.

Moreover, in acknowledgement of UK police guidelines which advise against feature-based composite construction when recall ability is limited (ACPO, 2009), a holistic composite system was employed. In the UK, police forces widely employ one of two holistic composite systems, either EFIT-V or EvoFIT. Whilst just over half of the UK's police forces employ EFIT-V (Solomon et al., 2012), EvoFIT has undergone more rigorous scientific testing in laboratory settings (see, e.g., Frowd et al., 2015; Frowd, 2017), and has also been employed in formal police field trials (Frowd, Pitchford, et al., 2012; Frowd et al., 2011). Consequently, comparatively more is known about the identification rates associated with EvoFIT.

Despite the rigorous testing of EvoFIT, however, to date there has been no research investigating the effect of constructing an EvoFIT composite on subsequent identification performance. The majority of previous research investigating the effect of composite construction has employed feature-based composite systems such as Identi-kit (Comish, 1987; Mauldin & Laughery, 1981; Yu & Geiselman, 1993), FACES (Kempen & Tredoux, 2012; Topp-Manriquez et al., 2014; Wells et al., 2005) or Photofit (Davies et al., 1978). Currently, only two studies have investigated the effect of holistic composite construction on identification performance, both of which employed EFIT-V (Davis et al., 2014, 2016). Within these studies, EFIT-V has produced mixed results, eliciting both a helpful effect (Davis et al., 2014), and a null effect (Davis et al., 2016) on subsequent identification performance. Considering the mixed results produced by EFIT-V, and the absence of any investigation into the effect of constructing an EvoFIT composite, it is arguable that any potential effects of holistic composite construction are not yet fully understood.

Experiment 1 therefore aimed to extend our understanding of the subsequent effects of composite construction by employing the holistic composite system EvoFIT, and utilising time delays that more closely approximate those experienced within applied settings. For this purpose, participants were assigned to either a *two-day* or *ten-week* time condition. Participants in the two-day condition were shown a short video clip that depicted a single target identity. Participants then proceeded immediately onto a cognitive interview, with composite constructors remaining to complete a facial composite. Control participants were not required to construct a facial composite and were dismissed after the cognitive interview. The recognition accuracy of both composite constructors and controls was then tested two days later using a target-present line-up. In the ten-week condition, participants viewed a video clip and then experienced a 24-48 hour time interval before participating in a cognitive interview and composite construction. This delay more closely reflects the time intervals used in applied settings (Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Frowd, Carson, Ness, Richardson, et al., 2005). Recognition accuracy was then assessed ten weeks later, which is more consistent with the delay that a real eyewitness may experience (see, e.g., Pike & Brace, 2002; Valentine et al., 2003).

In this instance, across both time intervals, recognition accuracy was comparable between composite constructors and controls. In the two-day condition, composite constructors made accurate identifications 90% of the time, with controls demonstrating 100% accuracy. Likewise, in the ten-week condition, recognition accuracy was at 60% among constructors, and at 65% for controls. Overall, these findings therefore suggest that the primary factor affecting recognition is retention interval, as significantly lower accuracy rates were elicited over ten-weeks, comparative to the two-day condition. The absence of a significant difference in recognition accuracy rates between constructors and controls supports previous research documenting a null effect (Davies et al., 1978; Davis et al., 2016; Yu & Geiselman, 1993). Nevertheless, this finding differs substantially to previous research indicating a harmful effect (Comish, 1987; Kempen & Tredoux, 2012; Topp-Manriquez et al., 2014; Wells et al., 2005).

Considering that all support for a harmful effect thus far has employed feature-based composite systems, it is possible that the *type* of composite system employed may contribute to the effect that construction incurs on subsequent identification performance. The wider face recognition literature has previously demonstrated that face recognition is a holistic process, whereby whole-face recognition is significantly easier than the recognition of individual facial features (Tanaka & Farah, 1993). Holistic composite systems are therefore thought to be more accurately aligned with the way in which faces are naturally encoded, as constructors only view whole-face arrays during the process of composite construction. The improved theoretical understanding that underpins holistic composite systems may therefore render any potential harmful effects of construction less likely.

Moreover, it has previously been suggested that feature-based composite systems produce less identifiable images than holistic composite systems (Frowd et al., 2015). It is therefore possible that previous research employing feature-based composite systems has produced a harmful effect as a result of composite constructors producing poor-quality facial composites. Indeed, a negative correlation has previously been found between composite quality and line-up performance (Davis et al., 2016; Wells et al., 2005). This correlation therefore suggests that the process of composite construction, in itself, may not be harmful. Instead, it is possible that the production of a poor-quality facial composite, that does not accurately resemble the original target, causes greater memorial interference and incurs a harmful effect on subsequent recognition tasks. In contrast, considering that EvoFIT is known to produce more recognisable images than those created on feature-based systems (Frowd et al., 2015), it is possible that Experiment 1 did not elicit a harmful effect due to increased composite quality. Overall, the findings of Experiment 1 suggest that holistic composite construction does not significantly affect subsequent identification performance and therefore supports the use of holistic composite systems in applied settings.

Having discovered in Chapter 2 that holistic composite construction elicits no subsequent effect on line-up performance, Chapter 3 investigated the possibility that some eyewitnesses may be better equipped for the task of composite construction than others. Considering that facial composites are only necessary when the perpetrator is unknown, the process of composite construction arguably relies on unfamiliar face recognition and face memory. Previous research in this domain has consistently shown that unfamiliar face recognition varies considerably between individuals (Bindemann, Avetisyan, et al., 2012; Bobak et al., 2016;

Burton et al., 2010; Russell et al., 2009), therefore suggesting that not all eyewitnesses may be equally equipped for the task. However, despite the broad distribution of abilities between individuals, face recognition is thought to be a relatively stable skill, whereby an individual's recognition performance remains relatively constant across different tasks (Bobak et al., 2016; Duchaine et al., 2007; Robertson et al., 2016; Russell et al., 2009), and across time (White et al., 2014; Wilmer, 2017).

Although the range and relative stability of face recognition has predominantly been studied theoretically, these findings may also have important implications for applied work, by providing an insight into eyewitness identification accuracy. Previous research has shown, for example, that an individual's performance on a standardised face test correlates with their ability to make an accurate eyewitness identification (Andersen et al., 2014; Bindemann, Brown, et al., 2012), even in distressing circumstances (Morgan et al., 2007). Chapter 3 therefore aimed to build on the notion that an individual's performance on standardised face tests relates to their aptitude as an eyewitness, and aimed to establish whether this relationship extends to the task of composite construction. Considering that holistic composite systems aim to closely align composite construction with the holistic processes involved in face recognition (Frowd, 2017), it is possible that an individual's face recognition ability also correlates with their ability to produce a good-quality composite, which accurately resembles a target face. Indeed, if individual differences in face recognition are related to composite quality, a means of further improving composite identification rates may be provided, by selecting only the most able witnesses for the task.

To investigate the role of individual differences in composite construction, Chapter 3 utilised two standardised recognition tests, comprising 40 trials of the 1-in-10 task (Bruce et al., 1999) and the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006).

Crucially, both of these tests are able to detect large individual differences in recognition performance (Bindemann, Brown, et al., 2012; Duchaine & Nakayama, 2006; Megreya & Burton, 2006; Russell et al., 2009), and have previously been used to assess eyewitness accuracy (Andersen et al., 2014; Bindemann, Brown, et al., 2012). Following completion of the 1-in-10 task and the CFMT, all participants viewed a photograph of an unfamiliar celebrity identity for one minute and subsequently created a facial composite of the target using the composite system EvoFIT. To establish the quality of the facial composites produced, a separate participant sample then rank-ordered the facial composites, based on their perceived level of similarity to the target identity.

A moderate correlation was found between a participants' overall performance on the 1-in-10 task and the CFMT, which is comparable to in size to other correlations across different tasks of face recognition (Fysh, 2018; Fysh & Bindemann, 2018; McCaffery et al., 2018; Robertson et al., 2017). However, individual performance on the 1-in-10 task and the CFMT did not correlate with a participant's average composite similarity ranking score, suggesting that composite construction may rely on different cognitive processes. The absence of a correlation between the two standardised recognition tests and composite quality converges with previous findings, which have found face recognition abilities to be unrelated to performance on a matching task utilising facial composites (Bate et al., 2018). Considering that both the current study and Bate et al.'s (2018) study found tasks involving facial composites to be unrelated to face recognition abilities, it is possible that the processing of facial composites relies on cognitive processes that currently remain undetected by recognition tests. For instance, although holistic composite systems aim to align construction with the holistic processing of faces through the presentation of whole-face arrays, composite construction may still differ substantially to other aspects of face recognition. The process involved when constructing an EvoFIT composite, for example, exposes constructors to very little variability between identities and requires constructors to repeatedly make selections based on minute differences (see Figure 5.1).

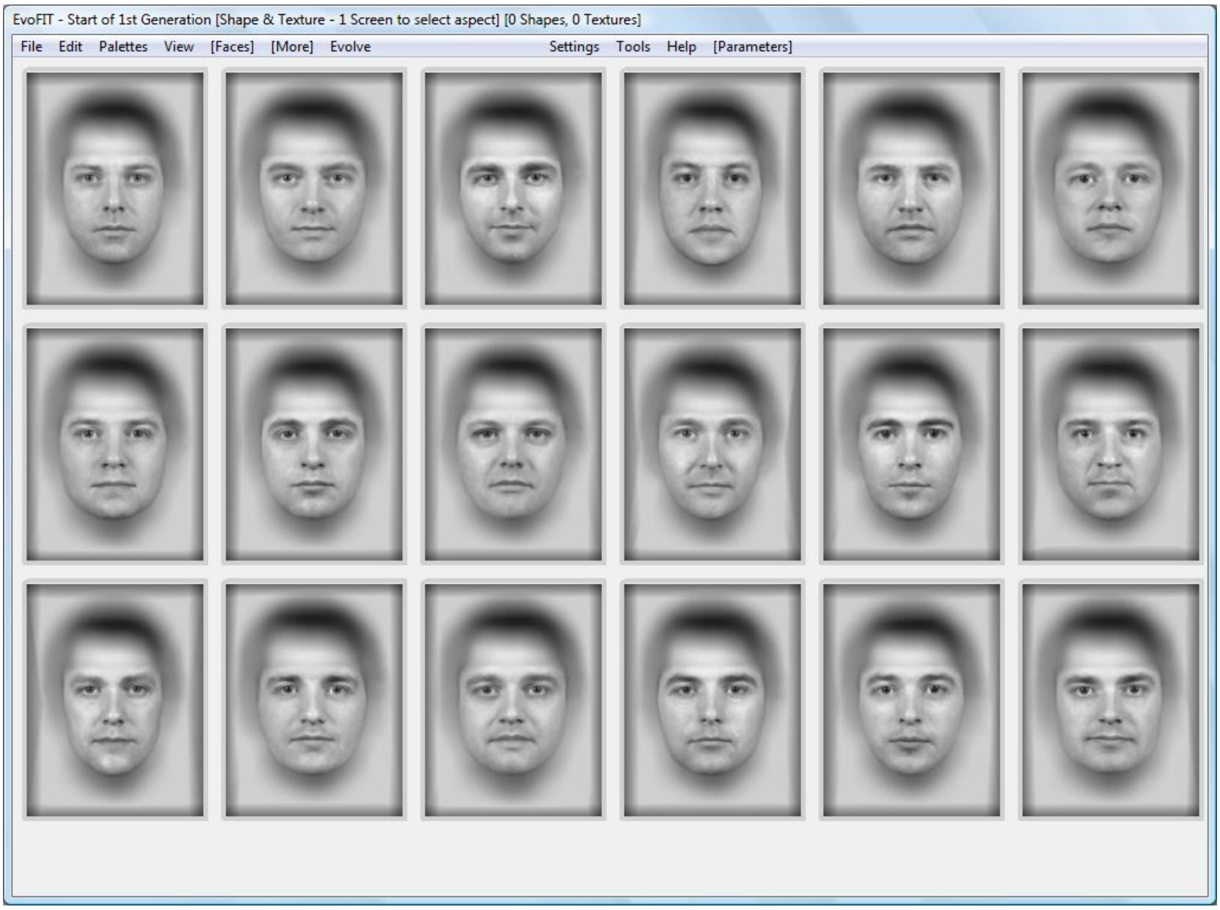

Figure 5.1. An example of an EvoFIT screen. Taken from Frowd (2012).

In contrast, natural face recognition is subject to considerably larger variability. For instance, multiple images of the same person can often be mistakenly perceived as different people by individuals who are unfamiliar with the target (Jenkins et al., 2011). The considerable *within*-person variation experienced in typical face recognition may therefore mean that differentiating between minute differences during composite construction more closely aligns the task with face perception. In contrast to face recognition tests, face perception tasks do not rely on face memory. These tasks may involve, for example, indicating whether a face pair is the same person (match) or different people (mismatch; see, e.g., Burton et al., 2010), or arranging facial images based on their perceived level of similarity to a target face (see, e.g., Duchaine et al., 2007). Indeed, although previous research has demonstrated a correlation between tasks of face perception and face recognition (Burton et al., 2010; Fysh, 2018), these correlations are far from perfect, indicating that the two tasks may utilise different cognitive processes. Overall, the findings of Experiment 2 suggest that composite construction is not a recognition task. It therefore still remains unclear which eyewitnesses are most suited for the task.

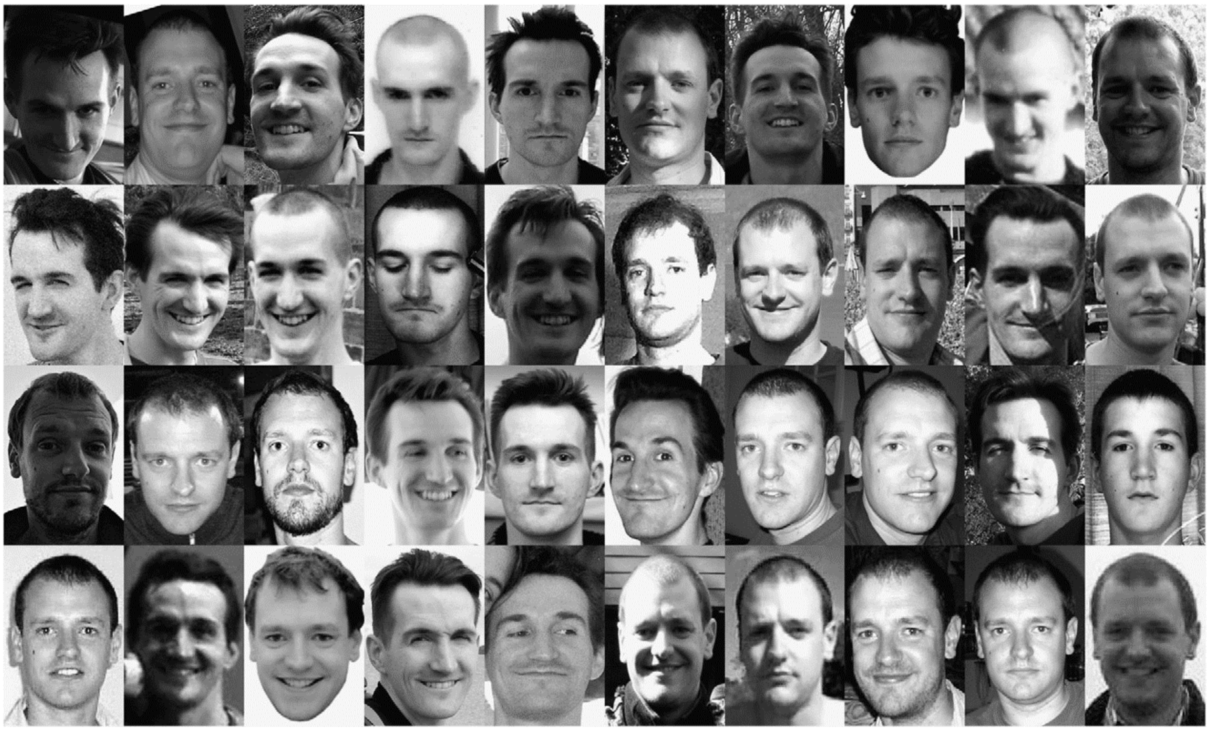

Figure 5.2. An example of within-person variation. Only two identities are shown here. Taken from Jenkins et al. (2011).

The final empirical chapter then sought to investigate another means of improving composite identification rates. In applied settings, when facial composites are released to the public, they are usually accompanied by contextual information, which typically provides additional details such as the perpetrator's clothing and so forth (see, e.g., Crimestoppers, 2018; Metropolitan Police, 2016). Whilst many psychological studies of composite naming have inadvertently acknowledged the role of context by, for example, revealing additional information about a target identity, such as their occupation (Frowd, Pitchford, et al., 2011; Frowd, Skelton, et al., 2011; Valentine et al., 2010), it currently remains unclear to what extent context facilitates composite recognition. Despite this, the wider face recognition literature has consistently shown that context aids face recognition (Kerr & Winograd, 1982; Krafka & Penrod, 1985; Mattarozzi et al., 2018).

In the context of familiar face recognition, for example, the worth of context is illustrated through theoretical accounts. The Interactive Activation and Competition model (IAC; Burton et al., 1990), for example, describes how familiar faces are recognised through the use of three central units comprising face recognition units (FRUs), person identity nodes (PINs) and semantic information (see Figure 5.3). In this instance, FRUs contain visual information about a face, whereas PINs allow access to semantic information, and are responsible for deciding whether a face is familiar or not. Using this architecture, the value of context is clear as semantic priming can result in a faster activation of PINs, and the subsequent recognition of a face (Bruce et al., 1993; Burton et al., 1990). Likewise, in unfamiliar face recognition, providing at least one piece of sematic information about a target identity has been found to facilitate subsequent recognition (Kerr & Winograd, 1982; Mattarozzi et al., 2018)

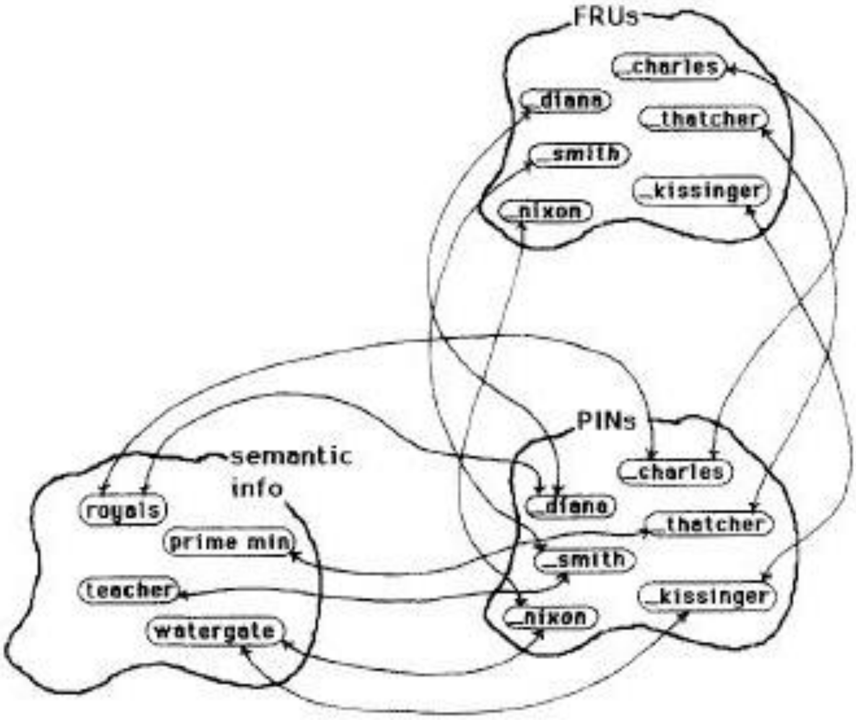

Figure 5.3. Taken from Burton et al. (1990). The central architecture of the IAC model.

Chapter 4 therefore sought to investigate whether the facilitating effect of context on face recognition extends to the recognition of facial composites. For this purpose, participants were asked to attempt to name facial composites created of 16 celebrity identities. The amount of contextual information that was displayed with these facial composites was varied systematically between three conditions comprising no-context, minimal context, and maximum context. In the no-context condition, participants sequentially viewed each of the 16 facial composites, and were asked to name the target identity. In this instance, no additional information was provided. In the minimal context condition, facial composites were accompanied by a single generic piece of semantic information about a person, such as their occupation. In the maximum context condition, three to four pieces of sematic information were provided that substantially reduced the number of possible identities (e.g. nationality, occupation, age and filmography). In addition, a context-only condition was employed, in which no facial composite was provided, to determine how effective context is for person identification on its own.

In this instance, a clear effect of context on recognition was found. In the no-context condition, correct naming was only achieved 5% of the time, which is comparable to spontaneous naming in other studies (e.g., Frowd et al., 2014; Valentine et al., 2010). However, in line with the wider face recognition literature, context was found to facilitate recognition (Kerr & Winograd, 1982; Krafka & Penrod, 1985; Mattarozzi et al., 2018). Naming rates increased from 5% to 10% when minimal context was provided, and increased again to over 55% when maximum context was displayed. However, naming rates were comparable across the maximum context and context-only conditions, suggesting that context, as opposed to composite images, may even present the primary factor for successful identification. It was therefore concluded that displaying a facial composite with at least three pieces of contextual information may aid composite identification rates.

The findings of this thesis have clear implications for the use of facial composites in applied settings. Historically, research has indicated that facial composites often depict a poor likeness of a target face and are consequently difficult to name (Davies et al., 2000; Frowd et al., 2007; Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005; Valentine et al., 2010). Although more recent holistic systems have improved identification rates substantially (Davis et al., 2017; Frowd, 2017; Frowd et al., 2012; Solomon et al., 2012), rather less research has investigated other means of further improving identification rates, aside from software developments. The current experiments support the use of holistic composite systems in applied settings, by indicating that the process of composite construction, in itself, elicits no subsequent effect on line-up performance (Experiment 1). It remains to be established, however, whether some eyewitnesses are better equipped for the task than others, based on individual differences (Experiment 2). This is an important avenue for future research given the broad distribution of abilities across individuals (e.g. Duchaine & Nakayama, 2006; Russell et al., 2009). Finally, strong effects of context on composite recognition were found, suggesting that displaying a facial composite with at least three pieces of contextual information may aid identification rates (Experiment 3).

5.2 Future research

In conclusion, this thesis investigated the use of facial composites in person identification and explored ways of improving holistic composite naming rates further. Considering the applied importance of this domain, it would be beneficial for future research to investigate whether the null effect of composite construction on line-up performance, found in Experiment 1, extends to other holistic systems. Although existing research employing an alternative holistic system, EFIT-V, has been found to elicit both a helpful effect and a null effect on subsequent identification performance, these findings were obtained over relatively short time intervals (Davis et al., 2014, 2016). Future research should therefore employ EFIT-V, which is also currently in police use, and test the effects of construction over forensically relevant delays. In turn, this will allow police forces to be more accurately advised on best-practice.

Moreover, Experiment 1 did not consider the effect of composite quality on identification performance. Considering that previous research has found a negative correlation between composite quality and line-up performance (Davis et al., 2016; Wells et al., 2005), it would have been beneficial to investigate whether misses (incorrect responses that the target is absent) and misidentifications (erroneous identifications of another line-up face as the target) were affected by the quality of a participant's composite image. In turn, this also provides an interesting direction for future work, whereby the effect of composite construction could be investigated systematically across different types of composite systems, and correlated with composite quality. Considering the conflicting findings within existing research, it would be beneficial to investigate the effect of composite construction on a between-subjects basis, employing composite systems from different generations. Such a design could employ, for example, a holistic system, a feature-based system and a sketch artist, to investigate whether composite quality significantly affects identification performance in each of these instances. These findings would therefore provide an insight into whether systems that generate less identifiable composites are more likely to produce a subsequent harmful effect on identification performance, thus potentially providing some insight into the effect associated with each type of composite system.

Further recommendations could also be made to police forces if future research can establish a means of selecting only the best eyewitnesses for the task of composite construction. The findings of Experiment 2 suggest that an eyewitness' face recognition abilities, and the quality of the facial composite they produce, are unrelated. It is therefore possible that real-world face processing tasks, such as composite construction, require more specific skills that currently remain undetected in standardised recognition tests. However, considering that the process of composite construction requires an eyewitness to differentiate between minute differences between identities (see Figure 5.1), an initial direction for future research could be to establish whether face perception abilities more closely align with composite quality. In contrast to recognition tests whereby subjects are required to identify a target from distractor faces, perception tests, such as the Cambridge Face Perception Test (CFPT; Duchaine et al. 2007), require subjects to arrange facial images according to their perceived level of similarity to a target face (see Figure 5.4). The ability to accurately differentiate between varying levels of similarity to a target may therefore be more accurately aligned with the process involved in composite construction. Investigating this notion may provide a means of improving composite identification rates further, by selecting only the most able eyewitnesses for the task. In turn, this may also save valuable police time, through the avoidance of constructing composites that are unlikely to be recognised.

Considering also that the true value of a composite is determined by correct naming rates, it would also be beneficial for future work to focus on the correlation between face perception abilities and composite naming. In Experiment 2, composite quality was determined by each facial composite's average similarity ranking score. This score was obtained by a group of 40 participants, who rank-ordered each facial composite, based on its perceived level of likeness to the target face. Considering that what constitutes a good likeness to a target face is based on an individual's level of familiarity with the target identity (Ritchie et al., 2018), it is possible that establishing composite quality in this manner yielded a noisy stimulus, affected by individual differences. Future work should therefore determine composite quality by naming rates, and investigate the correlation between an individual's face perception ability, and the rate their facial composite is correctly named.

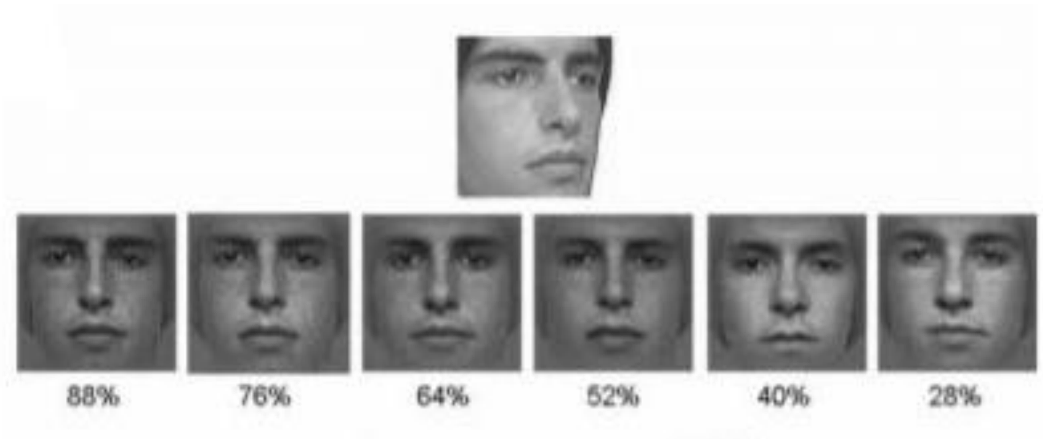

Figure 5.4. Taken from Duchaine et al. (2007). An example item from the Cambridge Face Perception Test. In this test, participants view six frontal face images and are required to sort these based on their level of similarity to the target.

Finally, considering that Experiment 3 indicated clear contextual effects on composite naming, another avenue for future research would be to investigate the *type* of contextual information that can be used for unfamiliar faces. Experiment 3 found that naming rates were substantially increased when three to four pieces of contextual information were provided, but the provision of context in this instance was dependant on the type of information that can be provided for celebrity identities (e.g. filmography). In order to build on the promising contextual effects demonstrated here, future research should accompany facial composites of unknown identities with varying short descriptions to establish which type of context is most likely to improve identification rates in applied settings.

Moreover, mean target naming was surprisingly low within the no-context (5%) and minimal context (10%) conditions in this instance. Recent work employing EvoFIT has yielded considerably higher naming rates, ranging from 24% to 74% (Frowd et al., 2013), therefore suggesting that differences in methodology may have negatively impacted naming rates. For example, Experiment 3 utilised facial composites which had been constructed with the target face in full view. This method of composite construction contrasts with the procedure used in applied settings, whereby a two-day delay is typically experienced between encoding and composite construction (Frowd, Carson, Ness, McQuiston-Surrett, et al., 2005). Although it is arguable that constructing a facial composite with the target in view should result in better quality composites, because memory factors are minimised (see, e.g., Estudillo & Bindemann, 2014), it is possible that this method of construction resulted in considerable focus being placed on individual facial features, thus interfering with EvoFIT's holistic approach to construction. To address this issue, future work could investigate whether facial composites constructed from memory generate similar contextual effects. In turn, this would more closely align the experimental process of composite construction with its use in applied settings, therefore allowing more definitive conclusions to be made regarding the provision of context in criminal investigations.

References

ACPO. (2009). Facial Identification Guidance 2009. Retrieved January 26, 2018, from http://library.college.police.uk/docs/acpo/facial-identification-guidance-2009.pdf

Andersen, S. M., Carlson, C. A., Carlson, M. A., & Gronlund, S. D. (2014). Individual differences predict eyewitness identification performance. *Personality and Individual Differences*, *60*, 36–40. https://doi.org/10.1016/j.paid.2013.12.011

Andrew, D. M., & Bird, C. (1938). A comparison of two new-type questions: Recall and recognition. *Journal of Educational Psychology*, *29*(3), 175–193. https://doi.org/10.1037/h0062394

Barkowitz, P., & Brigham, J. C. (1982). Recognition of faces: Own-race bias, incentive, and time delay. *Journal of Applied Social Psychology*, *12*(4), 255–268. https://doi.org/10.1111/j.1559-1816.1982.tb00863.x

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., … Richards, S. (2018). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications*, *3*(22). https://doi.org/10.1186/s41235-018-0116-5

Bindemann, M., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied*, *18*(3), 277–291. https://doi.org/10.1037/a0029635

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, *1*(2), 96–103. https://doi.org/10.1016/j.jarmac.2012.02.001

Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLoS ONE*, *11*(2). https://doi.org/10.1371/journal.pone.0148148

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. https://doi.org/10.1080/10683160500151159

Bruce, V., Dench, N., & Burton, M. (1993). Effects of distinctiveness, repetition and semantic priming on the recognition of face familiarity. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, *47*(1), 38–60. https://doi.org/10.1037/h0078773

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*(4), 339–360. https://doi.org/10.1037/1076-898X.5.4.339

Bruce, V., Ness, H., Hancock, P. J. ., Newman, C., & Rarity, J. (2002). Four heads are better than one: Combining face composites yields improvements in face likeness. *Journal of Applied Psychology*, *87*(5), 894–902. https://doi.org/10.1037//0021-9010.87.5

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*(3), 361–380. https://doi.org/10.1111/j.2044-8295.1990.tb02367.x

Burton, A. M., White, D., & McNeill, A. (2010). The glasgow face matching test. *Behavior Research Methods*, *42*(1), 286–291. https://doi.org/10.3758/BRM.42.1.286

Christie, D. F., & Ellis, H. D. (1981). Photofit constructions versus verbal descriptions of faces. *Journal of Applied Psychology*, *66*(3), 358–363. https://doi.org/10.1037/0021-9010.66.3.358

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. In A. M. Collins, E. E. Smith, A. M. Collins, & E. E. Smith (Eds.), *Readings in cognitive science: A perspective from psychology and artificial intelligence* (pp. 126–

136). San Mateo, CA, US: Morgan Kaufmann.

Comish, S. E. (1987). Recognition of facial stimuli following an intervening task involving the Identi-kit. *Journal of Applied Psychology*, *72*(3), 488–491. https://doi.org/10.1037/0021-9010.72.3.488

Crimestoppers. (2018). Most Wanted. Retrieved February 20, 2018, from https://crimestoppers-uk.org/most-wanted/?Page=2

Darling, S., Martin, D., Hellmann, J. H., & Memon, A. (2009). Some witnesses are better than others. *Personality and Individual Differences*, *47*(4), 369–373. https://doi-org.chain.kent.ac.uk/10.1016/j.paid.2009.04.010

Davies, G. M., Ellis, H., & Shepherd, J. (1978). Face identification: The influence of delay upon accuracy of Photofit construction. *Journal of Police Science and Administration*, *6*, 35–42. Retrieved from https://www.researchgate.net/publication/241195455

Davies, G., & Valentine, T. (2007). Facial composites: Forensic utility and psychological research. In R. L. Lindsay, D. F. Ross, J. D. Read, & R. L. Toglia (Eds.), *The handbook of eyewitness psychology, Vol II: Memory for people* (pp. 59–83). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

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. https://doi.org/10.1037/0021-9010.85.1.119

Davis, J. P., Gibson, S. J., & Solomon, C. J. (2017). Holistic facial composite systems: Implementation and evaluation. In M. Bindemann & A. M. Megreya (Ed.), *Face processing: Systems, disorders and cultural differences* (pp. 41–54). New York, NY: Nova Science Publishers.

Davis, J. P., Gibson, S., & Solomon, C. (2014). The positive influence of creating a holistic facial composite on video line-up identification. *Applied Cognitive Psychology*, *28*(5), 634–639. https://doi.org/10.1002/acp.3045

Davis, J. P., Thorniley, S., Gibson, S., & Solomon, C. (2016). Holistic facial composite construction and subsequent lineup identification accuracy: Comparing adults and children. *Journal of Psychology: Interdisciplinary and Applied*, *150*(1), 102–118. https://doi.org/10.1080/00223980.2015.1009867

Davis, J. P., Valentine, T., Memon, A., & Roberts, A. J. (2015). Identification on the street: a field comparison of police street identifications and video line-ups in England. *Psychology, Crime & Law*, *21*(1), 9–27. https://doi.org/10.1080/1068316X.2014.915322

Davis, R., Sutherland, N. S., & Judd, B. R. (1961). Information content in recognition and recall. *Journal of Experimental Psychology*, *61*(5), 422–429. https://doi.org/10.1037/h0044848

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. https://doi.org/10.1037/1076-898X.14.2.139

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, *24*(4), 419–430. https://doi.org/10.1080/02643290701380491

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

Ellis, H. D., Davies, G. M., & Shepherd, J. W. (1978). A critical examination of the Photofit system for recalling faces. *Ergonomics*, *21*(4), 297–307. https://doi.org/10.1080/00140137808931726

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. https://doi.org/10.1068/p080431

Estudillo, A. J., & Bindemann, M. (2014). Generalization across view in face memory and face matching. *I-Perception*, *5*, 589–601. https://doi.org/10.1068/i0669

Frowd, C. ., Park., J., McIntyre, A., Bruce, V., Pitchford, M., Fields, S., … Hancock, P. J. B. (2009). Evolving facial composite systems. *Forensic Update*, *97*, 25–32.

Frowd, C. D. (2012). Facial recall and computer composites. In C. Wilkinson & C. Rynn (Eds.), *Craniofacial Identification*. Cambridge University Press. https://doi.org/10.1017/CBO9781139049566

Frowd, C. D. (2017). Facial composite systems: Production of an identifiable face. In M. Bindemann & A. M. Megreya (Eds.), *Face processing: Systems, disorders and cultural differences* (pp. 55–86). New York, NY: Nova Science Publishers.

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

Frowd, C. D., Bruce, V., Ness, H., Bowie, L., Paterson, J., Thomson-Bogner, C., … Hancock, P. J. B. (2007). Parallel approaches to composite production: Interfaces that behave contrary to expectation. *Ergonomics*, *50*(4), 562–585. https://doi.org/10.1080/00140130601154855

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. https://doi.org/10.1348/135532504X15358

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. https://doi.org/10.1080/10683160310001634313

Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Perception*, *1*(1), 19–39. https://doi.org/10.1145/1008722.1008725

Frowd, C. D., Jones, S., Fodarella, C., Skelton, F., Fields, S., Williams, A., … Hancock, P. J. B. (2014). Configural and featural information in facial-composite images. *Science & Justice*, *54*, 215–227. https://doi.org/10.1016/j.scijus.2013.11.001

Frowd, C. D., McQuiston-Surrett, D., Anandaciva, S., Ireland, C. G., & Hancock, P. J. B. (2007). An evaluation of US systems for facial composite production. *Ergonomics*, *50*(12), 1987–1998. https://doi.org/10.1080/00140130701523611

Frowd, C. D., Pitchford, M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., … Hancock, P. J. B. (2011). The psychology of face construction: Giving evolution a helping hand. *Applied Cognitive Psychology*, *25*(2), 195–203. https://doi.org/10.1002/acp.1662

Frowd, C. D., Pitchford, M., Skelton, F., Petkovic, A., Prosser, C., & Coates, B. (2012). Catching even more offenders with EvoFIT facial composites. *Third International Conference on Emerging Security Technologies*, 20–26. https://doi.org/10.1109/EST.2012.26

Frowd, C. D., Skelton, F., Atherton, C., Pitchford, M., Hepton, G., Holden, L., … Hancock, P. J. B. (2012). Recovering faces from memory: The distracting influence of external facial features. *Journal of Experimental Psychology: Applied*, *18*(2), 224–238. https://doi.org/10.1037/a0027393

Frowd, C. D., Skelton, F. C., Atherton, C., & Hancock, P. J. B. (2012). Evolving an identifiable face of a criminal. *The Psychologist*, *25*(2), 116–119.

Frowd, C. D., Skelton, F. C., Butt, N., Hassan, A., Fields, S., & Hancock, P. J. B. (2011).

Familiarity effects in the construction of facial-composite images using modern software systems. *Ergonomics, 54*(12), 1147–1158. https://doi.org/10.1080/00140139.2011.623328

Frowd, C. D., Skelton, F., Hepton, G., Holden, L., Minahil, S., Pitchford, M., … Hancock, P. J. B. (2013). Whole-face procedures for recovering facial images from memory. *Science & Justice*, *53*, 89–97. https://doi.org/10.1016/j.scijus.2012.12.004

Frowd, C. D., White, D., Kemp, R. I., Jenkins, R., Nawaz, K., & Herold, K. (2014). Constructing faces from memory: The impact of image likeness and prototypical representations. *Journal of Forensic Practice*, *16*(4), 243–256. https://doi.org/10.1108/JFP-08-2013-0042

Frowd, C., Erickson, W. B., Lampinen, J. M., Skelton, F. C., Mcintyre, A. H., & Hancock, P. J. B. (2015). A decade of evolving composite techniques: Regression-and meta-analysis. *Journal Of Forensic Practice*, *17*(4), 319–334. https://doi.org/10.1108/JFP-08-2014-0025

Frowd, D., Hancock, P. J. B., Bruce, V., Skelton, F. C., Atherton, C., Nelson, L., … Frowd, C. D. (2011). Catching more offenders with Evofit facial composites: Lab research and police field trials. *Global Journal of Human Social Science*, *11*, 46–58.

Fysh, M. C. (2018). Individual differences in the detection, matching and memory of faces. *Cognitive Research: Principles and Implications*, *3*(20). https://doi.org/10.1186/s41235-018-0111-x

Fysh, M. C., & Bindemann, M. (2018). The Kent Face Matching Test. *British Journal of Psychology*, *109*, 219–231. https://doi.org/10.1111/bjop.12260

Gibson, S. J., Solomon, C. J., & Pallares-Bejarano, A. (2003). Synthesis of photographic quality facial composites using evolutionary algorithms. In R. Harvey & J. A. Bangham (Eds.), *Proceedings of the British Machine Vision Conference* (pp. 221–230).

https://doi.org/10.5244/C.17.23

Hancock, P. J. B., Bruce, V., & Mike Burton, A. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330–337. https://doi.org/10.1016/S1364-6613(00)01519-9

Hasel, L. E., & Wells, G. L. (2007). Catching the bad guy: Morphing composite faces helps. *Law and Human Behavior*, *31*(2), 193–207. https://doi.org/10.1007/s10979-006-9007-2

Innocence Project. (2016). Retrieved June 16, 2016, from http://www.innocenceproject.org/causes-wrongful-conviction/eyewitness-misidentification

Jenkins, F., & Davies, G. (1985). Contamination of facial memory through exposure to misleading composite pictures. *Journal of Applied Psychology*, *70*(1), 164–176. https://doi.org/10.1037/0021-9010.70.1.164

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323. https://doi.org/10.1016/j.cognition.2011.08.001

Kempen, K., & Tredoux, C. G. (2012). ‘Seeing is believing’: The effect of viewing and constructing a composite on identification performance. *South African Journal of Psychology*, *42*(3), 434–444. https://doi.org/10.1177/008124631204200315

Kerr, N. H., & Winograd, E. (1982). Effects of contextual elaboration on face recognition. *Memory & Cognition*, *10*(6), 603–609. https://doi.org/10.3758/BF03202443

Klatzky, R. L., Martin, G. L., & Kane, R. A. (1982). Semantic interpretation effects on memory for faces. *Memory & Cognition*, *10*(3), 195–206. https://doi.org/10.3758/BF03197630

Koehn, C. E., & Fisher, R. P. (1997). Constructing facial composites with the mac-a-mug pro system. *Psychology, Crime & Law*, *3*(3), 209–218.

https://doi.org/10.1080/10683169708410815

Kovera, M. B., Penrod, S. D., Pappas, C., & Thill, D. L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology*, *82*(2), 235–246. https://doi.org/10.1037/0021-9010.82.2.235

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality And Social Psychology*, *49*(1), 58–69. https://doi.org/10.1037/0022-3514.49.1.58

Kramer, R. S. S., Jenkins, R., Young, A. W., & Burton, A. M. (2017). Natural variability is essential to learning new faces. *Visual Cognition*, *25*(4–6), 470–476. https://doi.org/10.1080/13506285.2016.1242522

Mattarozzi, K., Colonnello, V., Russo, P. M., & Todorov, A. (2018). Person information facilitates memory for face identity. *Psychological Research*. https://doi.org/10.1007/s00426-018-1037-0

Mauldin, M. A., & Laughery, K. R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology*, *66*(3), 351–357. https://doi.org/10.1037/0021-9010.66.3.351

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*(21). https://doi.org/10.1186/s41235-018-0112-9

McKone, E., & Palermo, R. (2010). A strong role for nature in face recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11), 4795–4796. https://doi.org/10.1073/pnas.1000567107

McQuiston-Surrett, D., Topp, L. D., & Malpass, R. S. (2006). Use of facial composite systems in US law enforcement agencies. *Psychology, Crime & Law*, *12*(5), 505–517. https://doi.org/10.1080/10683160500254904

Megreya, A. M., & Bindemann, M. (2013). Individual differences in personality and face identification. *Journal of Cognitive Psychology*, *25*(1), 30–37. https://doi.org/10.1080/20445911.2012.739153

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*(4), 865–876.

Meissner, C. A., & Brigham, J. C. (2001a). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, *15*(6), 603–616. https://doi.org/10.1002/acp.728

Meissner, C. A., & Brigham, J. C. (2001b). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Memon, A., Meissner, C., & 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. https://doi.org/10.1037/a0020518

Metropolitan Police. (2016). E-fit released following sexual assaults. Retrieved February 20, 2018, from http://news.met.police.uk/news/e-fit-released-following-sexual-assaults-163154

Morgan, C. A., Hazlett, G., Baranoski, M., Doran, A., Southwick, S., & Loftus, E. (2007). Accuracy of eyewitness identification is significantly associated with performance on a standardized test of face recognition. *International Journal of Law and Psychiatry*, *30*, 213–223. https://doi.org/10.1016/j.ijlp.2007.03.005

Morgan, C. A., Wang, S., Mason, J., Southwick, S. M., Fox, P., Hazlett, G., … Greenfield, G. (2000). Hormone profiles in humans experiencing military survival training. *Biological Psychiatry*, *47*, 891−901. https://doi.org/10.1016/S0006-3223(99)00307-8

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, *128*(1), 56–63. https://doi.org/10.1016/j.cognition.2013.03.006

Office for National Statistics. (2017). Retrieved January 16, 2018, from https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/june2017

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied*, *19*(1), 55–71. https://doi.org/10.1037/a0031602

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pike, G., & Brace, N. (2002). *The visual identification of suspects: procedures and practice. (Briefing Note 2/02)*. A Publication of the Policing and Reducing Crime Unit.

Police And Criminal Evidence Act. (1984). Retrieved June 16, 2016, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253831/pace-code-d-2011.pdf

Postman, L., Jenkins, W. O., & Postman, D. L. (1948). An experimental comparison of active recall and recognition. *The American Journal of Psychology*, *61*, 511–519. https://doi.org/10.2307/1418315

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, *70*(5), 897–905. https://doi.org/10.1080/17470218.2015.1136656

Ritchie, K. L., Kramer, R. S. S., & Burton, A. M. (2018). What makes a face photo a ‘good likeness’? *Cognition*, *170*, 1–8. https://doi.org/10.1016/j.cognition.2017.09.001

Robertson, D. J., Jenkins, R., & Burton, A. M. (2017). Face detection dissociates from face identification. *Visual Cognition*, *25*, 740–748. https://doi.org/10.1080/13506285.2017.1327465

Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., Burton, A. M., & O'Toole, A. (2016). Face recognition by metropolitan police super-recognisers. *PLOS ONE*, *11*(2), e0150036. https://doi.org/10.1371/journal.pone.0150036

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

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. https://doi.org/10.1016/0010-0285(90)90003-M

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition*, *12*(4), 315–328.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, *100*(2), 139–156. https://doi.org/10.1037/0033-2909.100.2.139

Shepherd, J. W., & Ellis, H. D. (1973). The effect of attractiveness on recognition memory for faces. *American Journal of Psychology*, *86*, 627–633.

Solomon, C., Gibson, S., & Maylin, M. (2012). EFIT-V: Evolutionary algorithms and computer composites. In C. Wilkinson & C. Rynn (Eds.), *Craniofacial identification* (pp. 24–41). Cambridge University Press.

Sporer, S. L. (2001). The cross-race effect: Beyond recognition of faces in the laboratory. *Psychology, Public Policy, and Law*, *7*, 170–200. https://doi.org/10.1037//1076-8971.7.1.170

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.

https://doi.org/10.1080/14640749308401045

Topp-Manriquez, L. D., McQuiston, D., & Malpass, R. S. (2014). Facial composites and the misinformation effect: How composites distort memory. *Legal and Criminological Psychology*. https://doi.org/10.1111/lcrp.12054

Tredoux, C., Nunez, D., Oxtoby, O., & Prag, B. (2007). An evaluation of ID: An eigenface based construction system. *South African Computer Journal*, (37), 1–9.

Valentine, T., Davis, J. P., Thorner, K., Solomon, C., & Gibson, S. (2010). Evolving and combining facial composites: Between-witness and within-witness morphs compared. *Journal of Experimental Psychology: Applied*, *16*(1), 72–86. https://doi.org/10.1037/a0018801

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology*, *17*, 969–993. https://doi.org/10.1002/acp.939

Viper. (2009). Retrieved June 16, 2016, from http://www.viper.police.uk/index.html

Wells, G. L., Charman, S. D., & Olson, E. A. (2005). Building face composites can harm lineup identification performance. *Journal of Experimental Psychology*, *11*(3), 147–156. https://doi.org/10.1037/1076-898X.11.3.147

Wells, G. L., & Hasel, L. E. (2007). Facial composite production by eyewitnesses. *Current Directions in Psychological Science*, *16*(1), 6–10. https://doi.org/10.1111/j.1467-8721.2007.00465.x

Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory & Cognition*, *12*(4), 338–344. https://doi.org/10.3758/BF03198293

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. https://doi.org/10.1111/j.1529-1006.2006.00027.x

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.54.101601.145028

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PLoS ONE, 9*(8), e103510. https://doi.org/10.1371/journal.pone.0103510

Wilmer, J. B. (2017). Individual differences in face recognition: A decade of discovery. *Current Directions in Psychological Science*, *26*(3), 225–230. https://doi.org/10.1177/0963721417710693

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., … Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11), 5238–5241. https://doi.org/10.1073/pnas.0913053107

Yu, C., J., & Geiselman, R. E. (1993). Effects of constructing Identi-Kit composites on photospread identification performance. *Criminal Justice and Behavior*, *20*, 280–292. https://doi.org/10.1177/0093854893020003005

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., … Liu, J. (2010). Heritability of the Specific Cognitive Ability of Face Perception. *Current Biology*, *20*(2), 137–142. https://doi.org/10.1016/j.cub.2009.11.067