

FORENSIC FACIAL COMPOSITES

Please reference this book chapter as Frowd, C. D. (2021). Forensic Facial Composites. In A. M. Smith, M. P. Togliani, and J. M. Lampinen (Eds.) *Methods, Measures, and Theories in Eyewitness Identification Tasks* (pp. 34 - 64). New York, NY: Taylor and Francis.

There are many types of evidence available after a crime has been committed. Some evidence is physical, including fingerprints, footprints and DNA (e.g., Bradbury & Feist, 2005). Forensic officers attempt to recover information of this type from a crime scene for analysis. The hope, of course, is that the outcome will lead to identification the person(s) responsible. Success usually depends on recidivism: an offender has been convicted of a previous offence and his or her identifying information is available to be searched on a database of fingerprints, footprints, DNA, etc.

A similar situation applies to CCTV footage (e.g., Davies & Thasen, 2000). The hope now is that the offender has been caught on camera and a clear image of the face can be extracted. The image can then be compared against a database of known offenders for potential matches. It can also be circulated within a police force, or in the media, with the aim that someone who is familiar with the person will report him or her to the authorities. This use of evidence is psychological in nature due to the need for a human to recognise the face; it is considered by Richard Kemp and colleagues in Chapter 1 of this volume. A different situation also involving human recognition occurs when police officers have only seen a photograph of a wanted person, and so the face is unfamiliar, as discussed by Kara Moore and James Lampinen in Chapter 3.

As part of collecting all available evidence, the police will interview victims or bystanders who were present when the crime took place. Specific interviewing techniques have been developed for observers who may be able to provide an account of the crime and those involved (e.g., Fisher, 1995). Evidence collected in this way can be particularly important in the absence of the aforementioned evidence, or when recovering useful evidence may take a long time such that further crime occurs.

One (or more) of these observers might also be able to construct a composite image of the offender's *face*. This part of a person's appearance is particularly important for identification—although gait, voice, characteristic motion and possibly other human characteristics may facilitate recognition to some extent (as mentioned later). As with CCTV evidence, an image is circulated for identification. Again, the aim is that someone who is familiar with the person will name the face to police, providing an investigation with a potential suspect (e.g., Ellis & Shepherd, 1992). Subsequent police-work will collect evidence to assess whether or not the named person is likely to be responsible for the crime. Sometimes, a facial composite is constructed of someone with whom the police would like to make contact in order to eliminate him or her from the investigation, or to locate a potential witness, but composites are usually created of an offender, specifically a perpetrator of serious crime.

In this chapter, my aim is to assess facial composites as a reliable method of identification. Face-production systems have changed greatly over the years, as have methods of their deployment. Early 'feature' systems in the 1970s and 1980s are not reliable, nor are their computerised descendants, but the newer 'holistic' systems (esp. EvoFIT) can now create identifiable faces, for the first time providing a reliable means to identify offenders using this forensic technique. I also look to the future to consider promising techniques that may allow composites to be made even more effective.

State of the art

Over the years, practitioners have made use of many systems to interview an eyewitness, whether that person is a witness or a victim, to construct a composite face (e.g., Ellis & Shepherd, 1992; Fodarella, Kuivaniemi-Smith, Gawrylowicz, & Frowd, 2015). The earliest method involved artists, but "production" systems were later developed to allow (usually) police personnel with less artistic skill to interview eyewitnesses to create the face. These face systems were initially mechanical in nature, and then evolved into software programs. The general approach to create a face has changed considerably over time, initially with a focus on separate facial features—eyes, hair, nose, ears, etcetera—to implementations that involve a focus on the face as a whole, or on whole-facial regions.

One motivation for change is acceptable use of technology (e.g., ACPO, 2009), while another is production of more effective images (e.g., Frowd, Bruce, & Hancock, 2008). Feature systems, at least those using procedures developed to date, seem to have limited utility, but holistic systems (esp. one implementation) now create faces that are named (e.g., by police staff and the public) with very good accuracy. As will be seen, this favourable outcome for policing has emerged from developing systems alongside an initial witness interview and methods for triggering recognition of a finished composite.

Sketch. Production of sketched composites involves police officers and staff, or independent forensic artists. These professionals traditionally have training in portraiture, and work directly with a witness to draw the face on paper using pencils or crayons. A usual procedure is for artist and witness to consider distances between facial features, the *relational properties of the face* (e.g., Diamond & Carey, 1986), and then work to increase detail in the face. Artists may also make use of catalogues of facial features, to help witnesses identify or refine feature shapes to be drawn, as well as electronic media. Construction of sketches using remote interviewing procedures is also a viable option (Kuivaniemi-Smith, Nash, Brodie, Mahoney, & Rynn, 2014).

Readers interested in developing skill in sketching may appreciate the illustrated volume by forensic artist Karen Taylor (2000), or the procedural guide of Fodarella et al. (2015). Various methodologically-rigorous assessments have considered sketch production (e.g., Davies & Little, 1990; Frowd et al., 2005a, 2005b; Laughery & Fowler, 1980), several of which are reviewed below.

Mechanical systems. Two non-artist-based production systems emerged about 40 years ago (Ellis & Shepherd, 1992). The first was Identikit, popular in the US (Laughery & Fowler, 1980). This system contained sketch-like features (eyes, nose, mouth, etc.) printed on acetate slides. A technician would guide a witness through the process of selecting features, with chosen acetates stacked on top of one another to create the face. A further slide could be used for the technician to draw additional detail such as scars, marks and shading. The second system was Photofit, popular in the UK (Ellis, Shepherd,

& Davies, 1978). A technician would assemble selected features into a mechanical template to create the face. An acetate overlay can also be used to add further detail.

An intensive period of research ensued to assess the effectiveness of these systems (e.g., Ellis et al., 1975, 1978; Laughery & Fowler, 1980; Wells & Hryciw, 1984). The general approach involved recruiting participants, usually members of the public or university staff and students, as practice witnesses. Simulating a real investigation, these helpers would be shown a target face and construct a composite of it with help from a trained technician. The resulting faces would then be assessed for effectiveness by a second group of participants. These additional helpers would be asked to *match* the composites to a photograph of the target, or assess composites' goodness-of-fit or *likeness* (see also later section on *Proxies to naming*).

The findings did not bode well for mechanical systems. There is not space to discuss this impressive body of work, reviewed in Ellis and Shepherd (1992), except to mention two key findings. First, there were not sufficient features available to be able to create all faces (Davies, 1983). Part of the problem was that a given feature could not be resized. The problem was particularly relevant to Photofit as these features slotted into a template, and so could not be positioned. Second, Photofit contained 'dark lines' between one feature and another, as illustrated in Figure 1, the result of which interfered with recognition of the face itself (Ellis, Davies, & Shepherd, 1978).

Computerised 'feature' systems. Various issues were overcome with computerisation. For example, a large selection of facial features could now be stored in a software database for witnesses to select, and computer-graphics technology used to size and position features, and to blend features more acceptably on the face. The design of these software systems was also guided by emerging research that suggested that face perception is not a straightforward process.

It was becoming apparent that our experience of a face was interactive: perception of one part affected perception of another. More generally, research was converging on the notion that our experience of an upright face was *holistic* in nature. For example, recognition of a feature such as a

nose was more accurate when (i) it was embedded in a whole-face context rather than seen as an isolated part (Tanaka & Farah, 1993), and (ii) the spacing of features on the face—the relational information—was a better match to the target (Tanaka & Sengco, 1997). Even judgements about features were different in nature (uncorrelated) when made in isolation and in the context of an intact face (Davies & Christie, 1982). Further, recognition was facilitated when a face was encoded in its entirety (e.g., via attribution of character) rather than by its individual features (e.g., Wells & Hryciw, 1984). (See Chapter 6 by Vicki Bruce for further information on face perception.)

The implication is that facial features should be seen in a whole-face context. Thus, eyes should be more accurately selected by switching them in and out of a complete face. However, while convenient to show several pages of eyes, it was time-consuming and potentially taxing for a witness to view all example features in this way. The issue was overcome by inviting witnesses to give a description of the offender's face, information that could be used to limit the number of examples shown from the database of features. To obtain such a description, best practice made use of a cognitive interview (CI), originally designed by Ron Fisher, Ron Geiselman et al. (1985).

Cognitive interviewing contains techniques for facilitating recall of information. The CI has been subject to much development and readers may be interested in Milne and Bull (1999) and Wells, Memon, and Penrod (2007) for the technique in general, and Fodarella et al. (2015) and Frowd (2011) for CI as applied to face construction. In brief, interviewers should build rapport, and then ask witnesses to recall as much detail as possible, without guessing. The procedure, which for Photofit and Identikit had been straightforward, now required expertise and training (ACPO, 2000).

Composite systems and performance

Early assessments of performance. Several evaluations have been conducted on an early software system, Mac-A-Mug Pro. Koehn and Fisher (1997) followed real-world procedures where a trained interviewer guided participants through the process of describing and constructing an unfamiliar face

from memory. Other research involved participants creating the face by themselves (Kovera, Penrod, Pappas, & Thill, 1997), or employed an experienced operative to create composites with a target in-view (Cutler, Stocklein, & Penrod, 1988)—a procedure that should promote superior performance. Composite rating and matching-type tasks suggested that this system was unlikely to be effective.

Copying forensic practice. Other production systems emerged including Identikit 2000 and FACES, popular in the US, and E-FIT and PRO-fit, popular in the UK. My own research assessed systems such as these by carefully copying police use of composites (Frowd et al., 2005b). This design resulted in a ‘gold standard’ protocol involving design elements mentioned above. First, participants encode an unfamiliar face, the usual experience for witnesses (and victims) in a police investigation. Then, a trained interviewer uses cognitive-interviewing techniques to obtain a detailed description of the face from a witness, and takes each person through face construction as described by the product’s manufacturer—including, if possible, selection of features in the context of a complete face. The aim is construction of the best likeness possible, in an open-ended session, which if necessary also uses a paint program to add facial detail, improve the likeness of hair, etc. Crucially, and different to before, the resulting composites are shown to people who are *familiar* with the relevant identities to *name*.

Achieving a forensically valid result. To be of use to forensic practice, it is important to take into account that (i) witnesses differ in their ability to construct a face, (ii) one face may be easier to construct than another, and (iii) people name composites to a different level of accuracy. In particular, the protocol should consider *individual differences* in human ability—a fundamental aspect of human behaviour that I will return to later—and the *power* of an experiment to detect differences between conditions. While we usually do not know approximate differences between conditions without actually conducting an experiment, unless it is an attempted replication for example, detection of a so-called ‘medium’ effect size is a sensible objective (Cohen, 1988; Fritz, Morris, & Richler, 2012). In practice, this size is approximately 15% difference (or more) in the average (mean) values for composites between systems (or conditions)—for example, correct naming of 10% in one

system and 25% in another. This effect size (at least 15%) is therefore useful in the real world: it has practical significance. In order to achieve this objective, there should be at least 10 composites constructed per system (or condition) and at least 10 people who attempt to name the composites. Also, to be confident that results will generalise to other faces, there should be at least 10 different identities. An ideal design, then, as part of a between-participants design, would involve a set of 10 different identities constructed by different people in each system (condition), and a further 10 people for each system (condition) to name the composites.

It is worth mentioning that the above estimates are based on good experimental practice (e.g., incl. participants tested individually, clear instructions and use of randomisation procedures) and analyses using *t*-test or (for when there are more than two conditions in an experiment) Analysis of Variance (ANOVA)—for example studies, see Frowd, Skelton, Atherton, Pitchford, Bruce et al. (2012). Note that more effective statistics are now available, meaning that the power to detect true differences is greater, and such designs can involve fewer participants to achieve the same end result (for example studies, see Brown et al., 2018; Frowd et al., 2013; Skelton et al., in press). The above estimates also assume that correct naming responses to composites are not too infrequent or too often, as power is reduced for a ‘floor’ or a ‘ceiling’ effect, respectively. In such cases, it would be sensible to recruit more participants (and even involve more target identities) in order to maintain experimental power.

Proxies to naming. While the best assessment involves naming of composites, other methods have been deployed for decades (e.g., Ellis, Shepherd, & Davies, 1975; Laughery & Fowler, 1980) and still play a useful role today. These “proxies to naming” include ratings of likeness and matching-type tasks. The former normally involves assessors judging the match of each composite to a photograph of the intended identity. However, participants familiar with the target identities seem to rate more harshly, leading to low values and reduced experimental power, and so it is preferable to recruit participants for this task who are *unfamiliar* with the targets (e.g., Frowd, Nelson et al., 2012). The latter usually involves matching composites to photographs of the targets used in the experiment. As

such, the task would seem to be more objective than likeness ratings, although both often lead to similar conclusions (e.g., Davies, Milne, & Shepherd, 1983; Frowd, Bruce, Smith, & Hancock, 2008). Assessments (incl. naming) may also consider composites' internal or external features separately, providing information about the constructed accuracy of different regions of the face (e.g., Frowd, Skelton, Atherton, Pitchford, Hepton et al., 2012; Frowd, Skelton et al, 2011).

Assessment under favourable conditions. The first test of this protocol involved 10 celebrity targets (Frowd et al., 2005b). After confirming that the (randomly-selected) face was unknown, a participant-witness viewed the face for 60 seconds. Three to four hours later, the person worked with a trained interviewer to describe the face (using a CI) and construct a composite using one of five systems. These were E-FIT and PRO-fit software 'feature' systems, Photofit, a police forensic artist (who followed the same procedure but created the face by hand), and an early version of the EvoFIT 'holistic' system (described later). See Figure 1 for example composites.

Insert Figure 1 here

Composites from E-FIT and PRO-fit were equivalent, with a combined mean naming of 18.0% correct. Both were named better than Photofit, at 6.2%, a result which suggests that newer systems are more effective than old. Sketch composites were named at 9.2% correct. The assessment also revealed that E-FIT and sketch created faces that were more identifiable of distinctive compared to average-looking targets, reflecting an established result (e.g., Shapiro & Penrod, 1986). It was also apparent that, in spite of intentions to the contrary, proficiency in artistic ability was important, as illustrated in Figure 2, as was the need for good-quality training (see also Davies et al., 1983).

Insert Figure 2 here

In an analysis of five subsequent studies (two unpublished) using a similar design, including delay-to-construction of up to three-to-four hours and different types of target identity (incl. non-celebrity faces), Frowd et al. (2015) found naming to be similar, at 18.6% [$SE(N=5) = 4.9\%$] correct. Other assessments conducted at various research laboratories support this outcome (e.g., Brace, Pike, Allen,

& Kemp, 2006; Bruce, Ness, Hancock, Newman, & Rarity, 2002; Davis, Simmons, Sulley, Solomon, & Gibson, 2015). Overall, modern feature systems can construct somewhat-identifiable faces.

Increasing the retention interval. A different story emerges when a more realistic interval is used between target encoding and composite construction. In Frowd et al. (2005a), composites were created using the same careful procedures as described above, but two days after encoding. Composite naming from the modern feature systems was woeful: only two of the 10 PRO-fit composites were named correctly by anyone, mean correct of 1.3% overall, while none of the E-FIT composites were correctly named at all (out of 100 possible attempts). The project included the popular FACES system, but correct naming still remained low, at 3.2%. Only sketch was slightly more effective: Five composites were correctly named by at least one person, system mean at 8.1%.

This worrying finding has been replicated many times: usually, only about one in ten feature composites are correctly named, and mean naming is around 5% correct (e.g., Frowd, Bruce, Ness et al., 2007; Frowd & Fields, 2011; Frowd, McQuiston-Surrett, Anandaciva, Ireland, & Hancock, 2007; Frowd, Pitchford et al., 2010). In an attempt to be generalisable, studies have included (i) different types of targets (e.g., sports players, university staff, and characters from TV programmes), (ii) a retention interval between 24 and 48 hours, (iii) unpublished research (Frowd et al., 2015), to limit publication bias, and (iv) various feature systems. This outcome occurs despite following procedures likely to promote best results: with participant-witnesses aware that a composite would be required of them (and so encode the target intentionally), a reasonably-long encoding time (30 or 60 seconds), and interviewing by suitably-trained personnel (incl. a forensic artist with about 15 years experience of interviewing witnesses). These observations are supported by evidence from police that feature composites constructed in criminal investigations are rarely named (Frowd, Pitchford et al., 2012).

Understanding the problem

It has long been known that the feature-by-feature method used to construct these composites is at odds with the natural, global process of face recognition (e.g., Davies, Shepherd, & Ellis, 1978; Wells & Hryciw, 1984). While this observation hints at why identifiable faces are not produced routinely, it does not explain why composites are less effective after longer delays. In Ellis, Shepherd, and Davies (1980), participants encoded two unfamiliar faces, recalling each face immediately, and after a delay of one hour, one day or one week; at each interval, information recalled reduced significantly. Thus, information potentially useful for face construction is forgotten. This idea is illustrated in Frowd and Goodfellow (2017). Here, composites were created from descriptions of the face (used to locate features within the PRO-fit system) recalled immediately or 24 hours after encoding. Correct naming was significantly worse ($MD = 9\%$) for composites based on descriptions recalled after 24 hours.

Different regions of the face also seem to be important. Ellis, Shepherd, and Davies (1979) used facial photographs (not composites) and found that internal facial features (the area incl. eyes, brows, nose and mouth) were more effective for recognition than external facial features (hair, ears and face shape) for *familiar* identities. However, for composites created from memory, several projects report that composites' internal features are not more effective than external features (e.g., Frowd, Bruce, McIntyre, & Hancock, 2007; Frowd, Skelton, Butt, Hassan, & Fields, 2011). Thus, constructors would appear to have difficulty in creating the part of the face that is important for recognition by another person, internal features: instead, constructors seem rather adept at reproducing external features.

Creating more effective internal features

An important implication of this research is that techniques that promote construction of more accurate internal features should also give rise to more identifiable composites. This notion has actually led to methods that (i) improve face-recognition ability, (ii) exploit individual differences in construction ability, (iii) facilitate recognition of a finished composite, and (iv) use a different method (system) to create the face. I will now review each of these methods in turn.

Improving face recognition. Composite construction involves two main cognitive processes, face recall and face recognition. The cognitive interview administered at the start of face construction mainly concerns the former. As mentioned above, the CI contains specific techniques to improve a witness's face recall. While this information can be used to locate example features in a software system, or to sketch features on the page, face recall can be useful later, to be able to propose changes to enhance the likeness. However, selection of facial features is more relevant to face recognition. This is why modern systems should show features in a whole-face context, a procedure that promotes more identifiable composites (Skelton, Frowd, & Speers, 2015).

As such, improving a witness's face-recognition ability should lead to more accurate selection of facial features. This proposal has been assessed using two additional interviewing mnemonics, both 'holistic' in nature and designed to allow a face to be processed at a global level. Thus, after face recall, witnesses are asked to reflect on the perceived personality of the face, silently to themselves for one minute, and then via a series of judgements—health, masculinity, extroversion, etc. In practice, witnesses make seven such judgements on a three-point ordinal scale (low, medium and high). Face construction proceeds as normal. In Frowd, Bruce, Smith, and Hancock (2008), feature composites were correctly named at 8.6% when constructed using the usual face-recall CI but at an astonishing 41.2% following the enhanced interview: face-recall CI plus holistic mnemonics—together termed 'holistic' CI or *H-CI*. The H-CI led to more effective composites for the internal-features' region, as well as being more accurate by age, individual facial features, hair and overall likeness.

In a meta-analysis by Frowd et al. (2015) of seven studies, one unpublished, holistic mnemonics were effective for computerised feature and holistic systems, and for sketch composites when used in a specific way (see Frowd et al., 2015).

Individual differences in face construction. The holistic interviewing mnemonics provide a general benefit for face construction—that is, a *group* effect. However, as is the case with human performance, individual differences exist: some witnesses simply construct more identifiable faces

than others. Another approach, then, has explored how composites (of the same identity) from different witnesses might be combined to improve identification. While most crimes for which a composite is constructed involve only one witness, sometimes more than one person has seen the offender and so the approach has practical relevance.

One possibility might be to create a group composite. The potential benefit of individuals working in groups has been known for a long time (e.g., Hill, 1982; Miner, 1984; Osborn, 1957). For the current application, witnesses could be interviewed at the same time, coordinated by a practitioner, with the aim of reaching consensus for selection and placement of facial features. There is the potential for group discussion, to trigger memory, and to allow a composite to emerge. While there are issues around managing group dynamics, there is the potential that, in discussing details of an offender's face, witnesses may contaminate each other's memory. Indeed, there is considerable research showing the ease with which memory can be influenced (e.g., Hasel & Wells, 2007; Jenkins & Davies, 1985; Loftus, 1997). Consequently, and for good reason, some countries (e.g., UK) prohibit or limit (e.g., Australia) practices that may contaminate memory (ACPO, 2009; PACE Code D, 2017).

Vicki Bruce and colleagues (2002) assessed a much better method, where each witness constructed a composite independently. In one of their experiments, participants viewed a target identity and created the face using a feature system. Four composites were created likewise of the same identity, and these images were combined to create a '4-morph'; the procedure involves locating 'landmarks' on facial features, to allow software to produce a facial average. Correct naming of 4-morph composites ($M = 28\%$) was higher than the average of the best individual composite ($M = 22\%$), a finding supported by another of their experiments. The project also revealed that morphed composites were not made more effective by weighting constituent images with witness's own judgements of confidence. See Figure 1 for an example of the technique.

The authors explain that as witnesses worked independently, inaccuracies created in one composite should not be related to inaccuracies in another. Such uncorrelated information would then tend to

reduce when creating an image average. As a result of the research, UK police practice has been changed to allow construction of more than one composite (of the same identity) in criminal investigations (ACPO, 2000); the resulting morphed (averaged) composite, but none of its constituent images, can be released into the media for identification by members of the public.

The advantage of morphed composites has been replicated using a feature system (assessed by likeness ratings: Hasel & Wells, 2007; and by naming: e.g., Davis, Simmons, Sulley, Solomon, & Gibson, 2015; Frowd, Skelton, Atherton, Pitchford, Bruce et al. 2012) and found to extend to a holistic system (e.g., Valentine, Davis, Thorner, Solomon, & Gibson, 2010). The technique is also advantageous for combining faces created from another type of error-prone stimuli, forensic age progression of missing persons (Lampinen, Erickson, Frowd, & Mahoney, 2015).

Facilitating familiar face recognition. In spite of the considerable effort involved, a morphed composite is valuable when an offender has been seen by more than one observer. Other research has explored techniques when there is only one witness, and so only one composite is available—although see Ness (2001) for a novel way to produce a morphed composite from a single witness.

One technique from research capitalises on the notion that distinctive faces are easier to recognise (e.g., Shepherd, Gibling, & Ellis, 1991; Vokey & Read, 1992). The aim here is to *artificially* inflate the level of distinctiveness in the face and, in doing so, improve recognition. For example, if a composite has a large nose, exaggerating this feature should emphasise an important aspect of the face. As with morphed composites, facial landmarks are located, which are then exaggerated with respect to an average face, to create a caricature. Applied to photographic (non-composite) faces, an exaggeration is more recognisable than the original under certain circumstances—in particular, if the face is made harder to recognise by turning it into an outline drawing, or by brief presentation (e.g., Benson & Perrett, 1991, 1994; Lee, Byatt, & Rhodes, 2000). A composite is already difficult to recognise, and so emphasising distinctive aspects of the face may help to facilitate recognition.

Frowd, Bruce, Ross, McIntyre, and Hancock (2007) tested the technique with familiar-face composites exaggerated at 30% or 50% caricature. The procedure produced a small but non-significant increase in correct naming (cf. original composites). Further work revealed large individual differences in observers' preferred level of exaggeration for a given identity. Consequently, correct naming was found to improve when participants saw a composite changing by caricature, from negative to positive exaggeration, as illustrated in Figure 3. It was apparent that one frame in the sequence produced a better match to memory (cf. original), triggering recognition. Also, the sequence facilitated naming for composites produced from different systems, and when presented as a 'photospread' of frames or as an animation (e.g., <http://tiny.cc/animated-composite-1>). Frowd, Skelton, Atherton, Pitchford, Bruce et al. (2012) discovered that seeing all frames just once was sufficient, and that best performance ensued when both negative and positive parts were included.

Insert Figure 3 here

Other techniques have also been found to facilitate naming. Viewing a composite that has been stretched linearly in the vertical or horizontal direction seems to reduce the appearance of error and promote higher correct naming (Frowd et al., 2014), an effect which applies to both feature and holistic composites; benefit also extends to composites viewed side-on, a *perceptual* stretch, or when presented using a perspective 'backdrop' (see Figure 4). Similarly, naming is facilitated when a composite is misaligned horizontally, a manipulation that reduces the impact of inaccurate relational (configural) information (McIntyre, Hancock, Langton, & Frowd, 2016). Further, removing texture information in the face can increase accuracy (Frowd, Park et al., 2008), presumably by reducing overall error, as does adding sunglasses to composites with poor naming (e.g., McIntyre, Frowd, Bruce, & Hancock, 2010), concealing an inaccurate but clearly important part of the face. On a similar note, Brown et al. (2018) found it advantageous to add a hat and / or sunglasses to a composite. Such adornments conceal parts of a target's appearance that might have changed (e.g., hair) or had been constructed inaccurately (eyes). See Case Study B for an example of this technique in action.

Insert Figure 4 here

A more general issue. In spite of these developments, police procedures prohibit use of feature systems when witnesses struggle to recall all features of the face (ACPO, 2009). The rule was made to avoid witnesses from having to view too many facial features, yielding intractable composite sessions.

What was needed was a new way to produce the face, one that was not dependent on face recall.

Considerable effort has been spent on this challenge, the result of which are 'holistic' systems.

In the next part, I describe the general approach used for this type of system, and focus on the EvoFIT implementation that has developed a particularly effective algorithm for accessing memory. In fact, this algorithm has now been tested on a feature system, to allow more effective composites to be constructed (using this previous technology) in cases where witnesses have good recall.

Holistic face construction

Holistic systems require witnesses to choose from two or more faces (e.g., Frowd, Hancock, & Carson, 2004; Gibson, Solomon, Maylin, & Clark, 2009; Tredoux, Nunez, Oxtoby, & Prag, 2006). EFIT-V¹ (now called EFIT-6), for example, presents an array of nine faces, while EvoFIT presents 18 (see Figure 5).

Although the exact number shown is probably not important, the multiple-choice method of selection should lessen dependence on face recall, enabling witnesses to base decisions on *resemblance*. The result is a process that is arguably more closely aligned to face recognition (cf. recall), operating at a global or *holistic* level—hence the name given to this type of system.

Insert Figure 5 here

¹The latest version of this system, EFIT-6, does seem to involve a somewhat different approach to the one described here. The procedure for constructing a face is very flexible, led by the witnesses, with professionals tending to select facial features initially, and then making use of an evolving process at the end (cf. start of face construction) and/or adjusting the appearance of selected features.

Holistic systems usually share a common method of generating faces. Previously, systems took facial elements from different reference faces, which often caused difficulty in matching skin tone. Holistic types use internal features from facial photographs and apply statistics (usually Principal Components Analysis) to create a face model that is holistic in nature: changing one parameter has a global influence, perhaps making the face appear longer, older or healthier. Assigning random values to parameters creates plausible-looking faces that can be shown for selection. Then, parameters of selected items can be combined, with further selection, to allow a face to be “evolved”. Thus, an inherent property of these systems is that desired characteristics can be selected implicitly.

Various assessments have been conducted on holistic systems (e.g., Frowd et al., 2013; Tredoux et al., 2006; Valentine et al., 2010). Sadly, only one implementation (EvoFIT) seems to have been formally assessed with adults using the gold-standard protocol and a post-encoding delay of at least 24 hours. This is a shame, as arguably the best assessment for one implementation, EFIT-V, found naming at 20% correct, but the retention interval involved was unrealistically short (Valentine et al., 2010); other research appropriately suggests that composite likeness ratings decrease with increasing delay (Davis, Sulley, Solomon, & Gibson, 2011), but find performance is equivalent to a modern feature system (Davis, Simmons, Sulley, Solomon, & Gibson, 2015). Consequently, the following review focuses on EvoFIT. However, despite commonality among holistic systems, production of an identifiable face after a long delay turns out to require a radically different approach to the one described already.

Production of an identifiable face. In spite of considerable effort, early implementations of EvoFIT produced faces that other people rarely identified. In an early version, one which did appear to show promise, witnesses selected from arrays of smooth faces, items that changed by shape and position of facial features, and for overall head shape. Witnesses then selected from arrays of faces that changed by texture: colouring of eyes, nose, mouth and overall skin tone. After selecting the best matching combination of smooth and texture faces, the procedure was repeated with selected items combined to evolve a face, and then repeated likewise again. An evolved face could then be

enhanced using a “Shape tool”, to adjust size and position of features. In Frowd, Bruce, Ness et al. (2007), this procedure created composites, two days after target encoding, with mean naming of 11% correct—a system that seemed hardly worthwhile for policing. Previous attempts (e.g., those assessed in Frowd et al., 2005a, 2005b) were less effective.

Two developments made a remarkable difference, both improving the match of internal features and thus identification of the composite. First, after a witness had selected external features (hair, ears and neck), this region was blurred in the face arrays. The aim was to allow witnesses to more easily select internal features. See Figure 6 for example faces created in this way. The approach seemed better, but an evolved face still required considerable effort to improve the likeness, an issue that was addressed by the second development.

Insert Figure 6 here

To enhance the likeness of an evolved face more readily, ‘holistic’ scales were developed, each one altering the face as a whole (Frowd, Bruce, McIntyre et al., 2006). Examples include age, weight, attractiveness, health, masculinity (femininity) and extroversion. Scales were presented in a sequential order, with any changes made to the face carried forward to the next scale. Other scales were developed to alter shading of eyes, brows and mouth, and to add stubble, eye-bags, etc. Again, witnesses could use face recognition to judge changes that produced a better match. These tools have also proved useful to reduce artistic enhancement needed on a witness’s composite.

In Frowd, Pitchford et al. (2010), witnesses encoded an unfamiliar face for 60 seconds and created an EvoFIT composite using the gold-standard protocol after two days. When both developments were used together, the constructed composites were named with a fairly good mean, at 24.5% correct. This version of EvoFIT has now been assessed a further eight times, one unpublished, with a similar outcome and overall (unweighted) mean, at 23.0% [$SE(N=8) = 0.9\%$] (Frowd et al., 2015).

Subsequent research evaluated procedures with the aim of enhancing naming even further using the same design and long retention interval. First, Frowd, Nelson et al. (2012) found reliably higher

naming after witnesses underwent the face- plus holistic-recall (H-CI, $M = 39.4\%$) compared with face-recall CI alone ($M = 24.1\%$). Second, Frowd, Skelton, Atherton, Pitchford, Hepton et al. (2012) discovered that the level of external-features blurring used in face arrays was positively related to naming of an ensuing composite. Specifically, best performance emerged when blurring was set to maximum: then, only internal features were revealed (see Figure 6), and exterior features were chosen towards the end of the process, after holistic and shape tool use. These composites were named at 45.6% correct. Handling external features in this way forms part of the current operating procedure. Third, as mentioned above, naming of EvoFIT composites is facilitated if the face is physically or perceptually stretched (Frowd et al., 2014). Example EvoFITs are shown in Figure 7.

Insert Figure 7 here

Combining procedures. Frowd et al. (2013) assessed the combined effect of these three procedures. Participant-witnesses viewed a video of an unfamiliar target and constructed a face the following day using a face-recall CI or the H-CI, and when faces were presented with blurred or masked external features. The ensuing composites were named from the front and side-on. The research indicated benefit for each technique used separately, replicating the original research, but when techniques were combined, mean composite naming achieved an astonishing 73.8% correct.

Police field trials. In 2007, EvoFIT appeared suitable for policing. Research then suggested that best results involved use of a face-recall CI and external-features blurring, to produce composites after a two-day delay that could be named at around 25% correct (Frowd, Pitchford et al., 2010). System performance was assessed outside of the laboratory to ensure forensic training was suitable, check identification, elicit user feedback, etc. The first UK assessments were conducted by Lancashire and Derbyshire constabularies (Frowd, Hancock et al., 2010). Twenty-five police officers and staff were trained and supported during a six to 12-month trial period, and a total of 87 composites were constructed for a range of crimes, mainly serious. Composites were circulated among police officers, published in local newspapers, and posted on a wanted-persons' website using dynamic caricature.

Results revealed that, in 23.4% of cases, the person named from the composite was arrested. System enhancements ensued, promoting further trials by Devon and Cornwall Constabulary, and Romania, with arrests increasing to 40% (Frowd, Hancock et al., 2011). A further assessment by Humberside Constabulary involving [the improved procedure of] external-features' masking led to 60% arrests (Frowd, Pitchford et al., 2012). In this third assessment, 1 in 7 composites led to a name, an arrest and conviction of the person named. See *Case Study A* for an example of the system in use.

Assessment of usage. Impact in terms of use by forensic practitioners is appreciable. EvoFIT has been used for 31 police forces, UK and abroad, comprising over 3,000 composites constructed in mainly serious criminal cases between 2008 and 2013 (REF, 2014), and over 2,000 after this period to 2019 (Frowd, Portch, Killeen, Mullen, Martin, & Hancock, 2019; note that this reference includes a detailed assessment of modern composites involved in criminal investigations published in the media).

In the next part, I outline work that has applied effective techniques for EvoFIT to a feature system, in an attempt to improve this more traditional method of face construction. The research has led to a greater understanding of how people construct faces, and an unexpected enhancement for EvoFIT.

Back-propagating the findings. It seems fairly obvious that blurring external features should also benefit feature systems. When assessed with the usual face-recall interview, however, composite naming reduced. Frowd, Skelton, Battersby, and Fodarella (2018) suspected that the cause involved a mismatch of processing, similar to the established idea of Transfer-Appropriate Processing (Morris, Bransford, & Franks, 1977). Here, processing is unlikely to be optimal between a witness describing a face, an action leading to a focus of attention on both internal and external facial features, and face construction, with attention to internal features. However, this incongruence could be overcome using an H-CI, as that interview leads to focus on internal features (e.g., Frowd, Bruce et al., 2008). In an ensuing study, witnesses encoded an unfamiliar target and one-day later underwent a face-recall CI or an H-CI, and constructed the face as normal (whole face present throughout) or in parts. For the latter, internal features were presented and constructed first, then the opposite, external features,

and finally both internal and external parts were presented together, to allow further adjustment to be made. A third factor was composite naming viewed front-on and side-on.

Results revealed independent benefits for the H-CI and side-on naming. Method of face construction interacted with type of interview: correct naming was greatly reduced (cf. normal construction) when a face-recall CI was used with part-face construction ($M = 7.3\%$) but was greatly increased following an H-CI ($M = 79\%$); for an H-CI and part-face construction, mean composite naming from side-on was an astounding 87.1% correct. Thus, the work illustrates how very effective composites can be created using this method of face production, permissible (according to UK rules, ACPO, 2009) so long as a witness has good face recall. It also supports the theory that, for best performance, processing of the face needs to be consistent from one stage to another.

Further enhancement and alignment of stages in construction. Fodarella et al. (2017) investigated the impact of witnesses selecting for different *regions* in EvoFIT face arrays (Figure 5, right): (i) the lower half, including nose, mouth and chin, and (ii) the upper half, the region around the eyes, an area important for aspects of human communication including intention, emotion and identification (e.g., Shepherd, Davies, & Ellis, 1981). The work revealed different strategies in face selection, but significantly more identifiable composites were created when witnesses were instructed to select for the upper half (eye region) rather than the lower half or overall match. The importance of the eye region for identification is illustrated in Figure 8.

It turns out that focusing on this part-face (eye) region made the H-CI less effective (cf. face-recall CI) (Frowd & Logan, 2018). The mechanism is again thought to relate to a mismatch of processing between interview and construction. Frowd et al. (2018) segmented the H-CI, by following it with another cycle of character attribution, one where the witness was asked to focus on the eye region. This extra mnemonic facilitated composite naming, allowing character attribution to be effective again. The enhanced (now three-part) interview forms part of the standard protocol for practitioners.

Witness and target factors. Face construction has been assessed in relation to vulnerable witnesses and victims (Davis, Thorniley, Gibson, & Solomon, 2015; Fodarella, Brown, Lewis, & Frowd, 2015; Gawrylowicz, Gabbert, Carson, Lindsay, & Hancock, 2012; Paine, Pike, Brace, & Westcott, 2008), when large differences in age exist between constructor and target (Fodarella, Brown, Lewis, & Frowd, 2015), and when stress (Hancock, Burke, & Frowd, 2011) or distraction (Marsh et al., 2015, 2016) are present at encoding. Research has improved witness performance using a mindfulness (anxiety-reducing) technique (Martin, Hancock, & Frowd, 2017), and has explored the impact of properties of a target: facial distinctiveness (Frowd et al., 2005b) and attractiveness (Richardson et al., in prep).

Carry-over effects. Extensive face recall can impinge on face recognition (e.g., Alogna et al., 2014), as might close scrutiny of facial features during face construction, and so there has been considerable interest in a constructor's ability to accurately select a target from a line-up (identity parade). Findings have been mixed, arguably due to differences in production systems and methods for assessing witness performance. A meta-analysis of eight studies for the now archaic mechanical systems indicates a moderate positive effect on identification (Meissner & Brigham, 2001), while some concern has been raised about sketch construction in one study (Hall, 1976). More recent research involving the FACES feature system indicates a sizeable negative effect on identification for both face constructor and another person who has merely viewed the resulting composite (Kempen & Tredoux, 2012; Topp-Manriquez, McQuiston-Surrett, & Malpass 2016; Wells, Charman, & Olson, 2005). For holistic systems, effects range from null (Davis, Thorniley, Gibson, & Solomon, 2015; Turner et al., 2016) to positive (Davis, Gibson, & Solomon, 2014).

Chapter summary

A great deal of effort has been dedicated to assessment and development of composite systems, and for good reason: an effective composite can provide an all-important name to kick-start a criminal investigation. A composite that elicits a name can, given initial supporting evidence, provide a focus

for the police to determine whether or not the person named is to be treated as a suspect for the offence in question.

Methods for constructing faces have changed greatly over the last four decades. While artist's sketches remain an important forensic tool in many countries, production systems have moved away from mechanical types where facial elements are assembled as jigsaw pieces or printed on acetate slides. Computerisation has greatly improved the process, to allow systems to have a broader choice of features and to create more realistic-looking faces. Witnesses are now asked to recall the appearance of the offender's face and facial features can be selected in the context of a whole face. However, using these traditional systems with witnesses after a long retention interval usually result in composites that are rarely identified, limiting their utility.

Various initiatives have facilitated performance, in particular those that have made composites' internal-features region more effective. Novel techniques have improved a witness's ability to select facial features (via character attribution); recruited more memory, to allow production of a less error-prone representation (a morphed composite); and presented observers with enhanced versions of a composite, to trigger recognition. However, many witnesses are denied the opportunity of creating a composite, as they cannot recall the offender's face in detail.

Holistic systems were developed to rectify this situation. Witnesses now repeatedly select from screens of faces, with choices combined, to allow a composite to be evolved. The chapter's focus was on the much-researched EvoFIT implementation. De-emphasis or (better) removal of external features is valuable for production of an identifiable face, to allow witnesses not to be distracted by the exterior region. A character-based interview was also found to facilitate witness performance, as were post-production methods to help observers name a witness's composite: dynamic caricature, stretching and side-view. Further research revealed the appreciable value of combining these techniques, to produce composites with high mean correct naming, at over 70%.

De-emphasis of external features was then applied to a feature system. While unsuccessful when a face-recall CI had been administered, the combined protocol of H-CI, construction first by internal features and naming of the face from a side-view promoted very effective composites: mean correct naming at over 80%. The underlying mechanism is arguably due to alignment of interview, construction and naming. This mechanism has also been applied to a recent development of EvoFIT involving selection of the eye region, with the finding that H-CI is more effective when witnesses are requested to repeat character attribution for this more specific region of the face.

The future

What then is the future of facial composites? It is clear that these forensic images are much more effective for identifying suspects than ever before. The computerised feature systems, while usually ineffective when deployed as originally designed, have huge potential. Arguably their Achilles Heel relates to witnesses' poor face recall, and recent work is exploring how to overcome this issue. In particular, there would appear to be mileage in recalling the face earlier. Frowd, Bunce, Madin, and Brown (2016) asked witnesses to recall the face by themselves three-to-four hours after target encoding, with face construction the following day. This protocol led to more effective composites (cf. no self-administered recall). The team are attempting to replicate the effect and see whether it extends to different time intervals for self-recall. If their endeavours are successful, the consequence is exciting, since the change to procedure would be minor, but the benefit appreciable. One would anticipate that self-administered recall would also facilitate composites drawn by a forensic artist.

The notion of witnesses working by themselves is intriguing. Issue of collusion aside, there is no real reason why a witness should not construct an entire composite by him or herself. Many crimes go unsolved, but creating a composite in cases where previously one would not be produced could have a dramatic impact on volume (minor) crime. In fact, an effective self-administered protocol is clearly valuable in situations where police resources are limited, or where social-distancing measures are

necessary for practitioners. With these motivations in mind, we have been assessing self-administered procedures, both for standalone (non-Internet) and online applications. Using a simple face-recall CI, Martin et al. (2018) found that composites were self-constructed after a long delay with naming of about 25% correct. This initial work highlighted the importance of ergonomic processes and, if an online app is used, the need to limit Internet delays. Recent research using a more effective version of EvoFIT and a further simplified interview has allowed participants to create composites themselves that are named at around 45% correct (www.uclan.ac.uk/research/explore/projects/self-administered-procedures-facial-composites.php), of clear forensic importance. Similarly, Kuivaniemi-Smith et al. (2014) has found that a sketch constructed over a communication channel is an effective method of face production.

Research is also assessing the impact of witnesses recalling the environment (context) in which a target had been seen, with overall benefit found for sketch, EvoFIT and PRO-fit (Fodarella, Marsh, Chu, Wilcock, & Frowd, 2018; Kuivaniemi-Smith, Richardson, Nash, Uther, & Frowd, 2018). Further, research is assessing the nature of faces shown in EvoFIT arrays, to make the task potentially less taxing for witnesses: (i) in spite of reducing the search space for evolving a face, emerging data suggest that presenting fewer arrays actually allows constructors to create more effective composites, and (ii) while viewing just internal features is effective, the overall likeness sometimes reduces when external features are then added (presumably due to the interactive nature of face perception), and so it might be preferable both to blur external features to some extent at the start and to give specific instructions for face selection (e.g., focus on the eye region). Other research is exploring the novel idea of using holistic scales to alter individual facial features (cf. whole face), thus allowing greater refinement of the face towards the end of the process than had been possible previously. Furthermore, adding person-specific motion to a witness's composite also appears to be useful to facilitate recognition; the approach is similar to dynamic caricature but using motion that is characteristic of the target identity (Lander, Portch, Frowd, & Cootes, 2017).

As a final comment, it is a pleasure to be able to report that the combination of appropriate research methodology and application to theory has allowed composites to evolve as an effective tool for policing. There is now an international community of researchers working on the common goal of facilitating composites' performance. The future would appear to be promising for this forensic method of identification.

Case Studies

There have been many cases where EvoFIT has been valuable to law enforcement to help solve crime. In this section, one case is described that led to identification and subsequent conviction of a dangerous person, a serial rapist, while in another, EvoFIT helped to locate a murder suspect on the run. Further examples can be found in Frowd, Hancock et al. (2010) and (2011), Frowd, Pitchford et al. (2012) and Frowd et al. (2019).

Case Study A: Catching a serial rapist

As reported widely in the media, Operation Hatton involves serious sexual assaults committed by a young adult around 2009 and 2010. Victims were female, particularly young students in the south part of Manchester. Offences occurred in the early hours, were violent, and the offender was clearly callous. In one crime, a weapon was involved, while in another the offender actually reassured a victim's father that his daughter was safe and being brought home. CCTV reveals the predatory nature of these awful offences. In spite of a large-scale police campaign, including use of a composite image, the offender remained at large.

In the most recently reported case, seven days after the offence, the victim constructed an EvoFIT composite. The procedure involved the normal face-recall interview, with EvoFIT presenting internal features in face arrays and holistic tools; the victim proceeded through each screen, selecting faces for overall best match. External features were chosen towards the end of the process, which were lighted, to better match the victim's memory of the rapist. The resulting image is shown in Figure 8.

Insert Figure 8 here

Greater Manchester Police released the composite in a public appeal for information, also using dynamic caricature (<http://tiny.cc/animated-composite-1>). The outcome was a huge public response, with the name *Asim Javed* reported frequently. This person was an employee of a fast-food outlet in

the Chorlton area of Manchester. Javed was arrested and the investigation proceeded to gather evidence. During interview, Javed confessed that he would have continued to offend had he not been stopped. He received an indefinite prison sentence.

Case Study B: Catching a murder suspect on the run

The use of EvoFIT in a complex and lengthy criminal investigation in Wales provides another illustration of its usefulness. As this system can depict offenders who disguise their appearance, as well as to overcome inaccuracies present in a constructed composite (Brown et al., 2018), EvoFIT was deployed in a widely-advertised investigation to locate murder suspect Wayne David Tidy, a person who had been at large for several weeks (Figure 9, far left). An EvoFIT composite of Mr Tidy was constructed (second from left) and a range of disguises from the system were applied to the face. These images were used for intelligence, to indicate his likely appearance. An officer from Dyfed Powys Police who was involved in the investigation reported that “EvoFit images were very helpful as it allowed the officers who finally arrested him to envisage TIDY in the numerous guises”. On arrest (far right), the suspect had grown a beard similar to the centre-right depiction. He was convicted of the murder of Simon Clark and sentenced to life, minimum of 24 years. Further details about this case can be found in a press release from Dyfed-Powys Police (2019).

Insert Figure 9 here

Figure 1. Composites produced in Frowd et al. (2005b). Different participants constructed composites using, from left to right, E-FIT, PRO-fit, sketch and Photofit. Can you guess the well-known person depicted here? The image on the far right is the average ('morphed') composite containing the sketch and the other modern composites of this person: this face should be a more effective cue to identity. The answer is mentioned at the end of the chapter.



Figure 2. The potential value of artistic enhancement. In this example, a practitioner selected features using the PRO-fit feature system (left) and then added artwork using the in-built painting tools (right), to improve recognisability of the celebrity face. (See end for identity.)

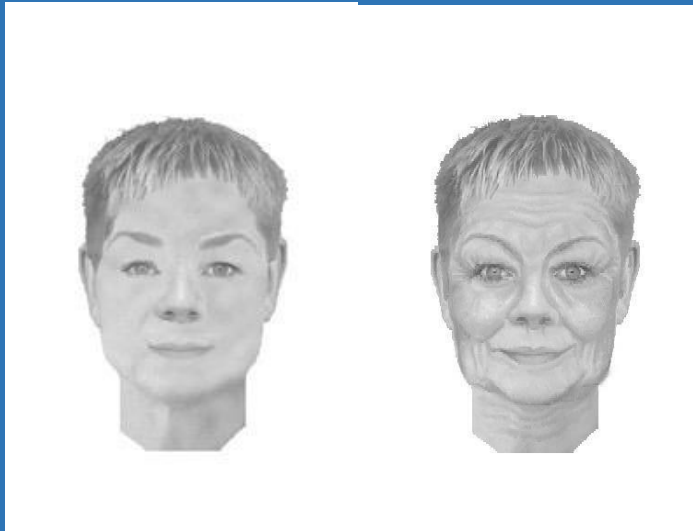


Figure 3. In this example, a composite of a familiar identity was constructed from memory during practitioner training. The shape information in the face was then exaggerated and de-emphasised with respect to an old-age male face, to produce the sequence shown (from left to right) at -50%, -25%, 0%, 25% and 50%. Values less than zero produce negative or anti-caricature (a face more like the average), while values greater than zero produce positive caricature (an exaggeration); the centre image (0%) is the original. In practice, members of the public are usually shown a dynamic sequence to identify, comprising an animated GIF of 21 frames played for a period of 6 seconds.



Figure 4. Recognition can be enhanced (cf. original, right) when a finished composite is seen stretched (left) or (more conveniently) when seen as part of perspective backdrop (centre).

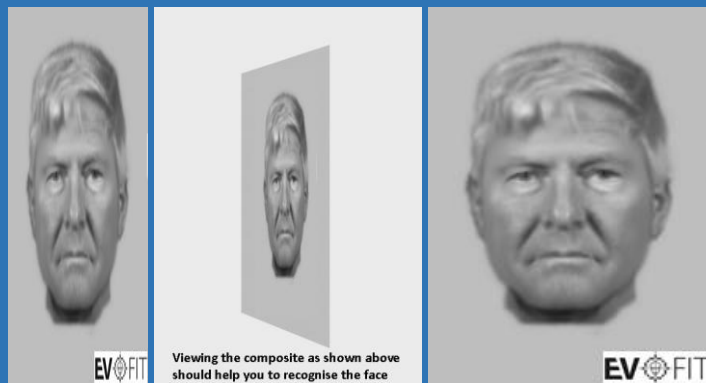


Figure 5. Example face array. An early (and generally ineffective) EvoFIT interface presented face arrays like the one shown here, with external features present, for a witness to select best matches. It was used for the first time in Operation Mallard (Frowd, Bruce, Storås, Spick, & Hancock, 2006).

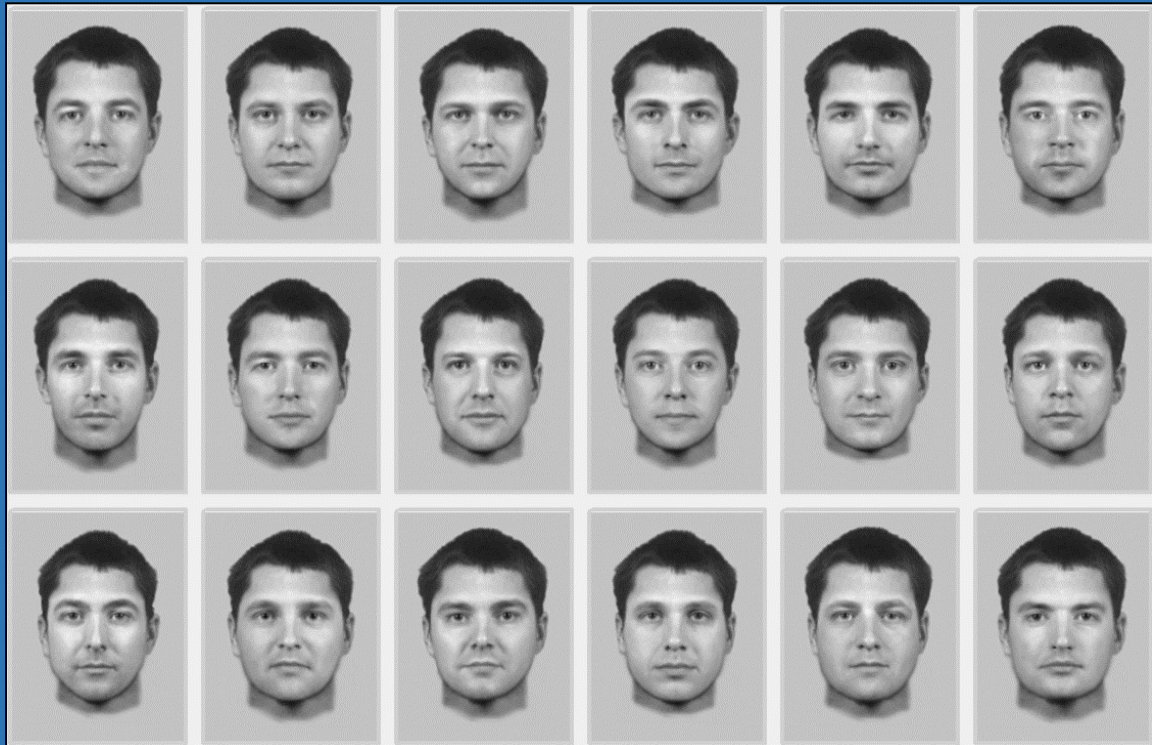


Figure 6. On the left is an example EvoFIT array with external-features blurred, and right, containing internal features only (with external features added to the face towards the end of the process).

Readers should find it easier to perceive internal features on the left array rather than those presented in Figure 5; perception of internal features should be even easier without external features (right).

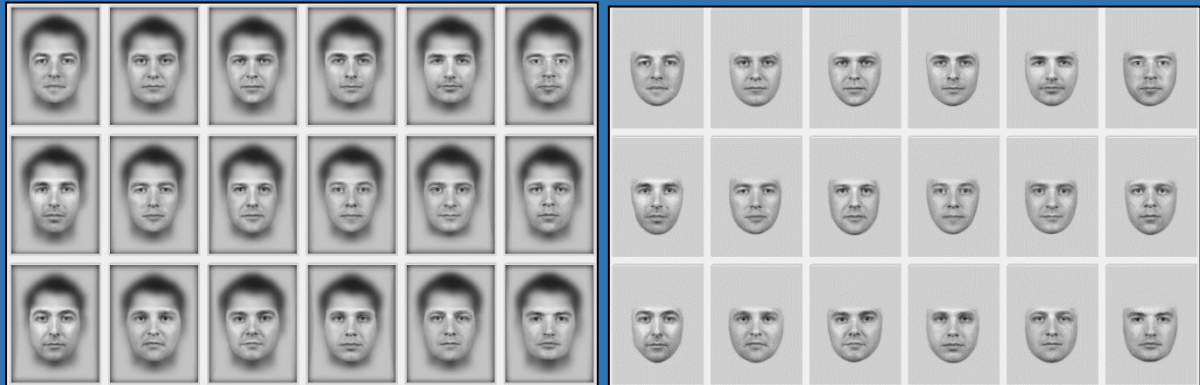


Figure 7. The following composites were constructed from memory of celebrity faces mainly as part of EvoFIT training. Can you guess the identities depicted? You might notice that the likeness does not have to be an exact match for the face to be identifiable; also, that some faces seem easier to recognise than others. Answers are listed at the end of the chapter.



Figure 8. The EvoFIT composite constructed in Operation Hatton, left, and a picture of the person who was convicted in this police investigation.



Figure 9. Catching a murder suspect (far left) at large. In this case, a composite and a range of facial depictions were prepared to brief officers as to this person's likely appearance. A photograph of the suspect at arrest is shown on the far right. See chapter text for more details.



Answers

It is not the author's intention to suggest that celebrities depicted in this chapter are in any way responsible for committing crime. Their inclusion is purely for illustrative purposes. It is perhaps worth mentioning here that if any of these celebrities are not known, a simple search on the Internet should reveal the relevant person's face.

Figure 1. Retired US tennis player Andre Agassi.

Figure 2. British actress Dame Judy Dench.

Figures 3 and 4. US President Donald Trump.

Figure 7. The celebrity composites are of music producer Simon Cowell, actor David Jason, music artist Will.i.am, former US President George W. Bush, English TV presenter Anthony (Ant) McPartlin, musician Robbie Williams, actress Whoopi Goldberg, physicist Albert Einstein, actress Julia Stiles and musician Noel Gallagher (former band member of Oasis).

ACPO (2000). *National Working Practices in Facial Imaging*. UK Association of Chief Police Officers Working Group. Unpublished document.

ACPO (2009). *Facial Identification Guidance*. Produced by the National Policing Improvement Agency. Unpublished document.

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahnik, S., Birch, S., Birt, A. R., et al. (2014). Registered replication report: Schooler & Engstler-Schooler (1990). *Perspectives on Psychological Science*, *9*, 556-578.

Benson, P. J., & Perrett, D. I. (1994). Visual processing of facial distinctiveness. *Perception*, *23*, 75-93.

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

Brace, N., Pike, G., & Kemp, R. (2000). Investigating E-FIT using famous faces. In A. Czerederecka, T. Jaskiewicz-Obydzinska & J. Wojcikiewicz (Eds.). *Forensic Psychology and Law* (pp. 272-276). Krakow: Institute of Forensic Research Publishers.

Bradbury, S. A., & Feist, A. (2005). The use of forensic science in volume crime investigations: A review of the research literature. Research Development and Statistics Directorate, Home Office, *Online Report 43/05*.

Brown, C., Portch, E., Skelton, F. C., Fodarella, C., Kuivaniemi-Smith, H., Herold, K., Hancock, P. J. B., & Frowd, C. D. (2018). The impact of external facial features on the construction of facial composites. *Ergonomics*, *62*, 575-592.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY.

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

Davies, G. M. (1983). Forensic face recall: The role of visual and verbal information. In S.M.A. Lloyd-Bostock and B.R. Clifford (Eds.). *Evaluating witness evidence* (pp. 103-123). Chichester: Wiley.

Davies, G. M., & Christie, D. (1982). Face recall: An examination of some factors limiting composite production accuracy. *Journal of Applied Psychology, 67*, 103-109.

Davies, G. M., Milne, A., & Shepherd, J. W. (1983). Searching for operator skills in face composite reproduction. *Journal of Police Science and Administration, 11*, 405-9.

Davies, G. M., Shepherd, J. W., & Ellis, H. D. (1978). Remembering faces: Acknowledging our limitations. *Journal of Forensic Science, 18*, 19-24.

Davies, G. M., & Thasen, S. (2000). Closed-circuit television: How effective an identification aid? *British Journal of Psychology, 91*, 411-426.

Davis, J. P., Simmons, S., Sulley, L., Solomon, C., & Gibson, S. (2015). An Evaluation of post-production facial composite enhancement techniques, *Journal of Forensic Practice*, 17, 307-318.

Davis, J. P., Sulley, L. Solomon, C., & Gibson, S. (2011). A comparison of individual and morphed facial composites created using different systems. In G. Howells, K. Sirlantzis, A. Stoica, T. Huntsberger and A. T. Arslan (Eds.) 2010 IEEE International Conference on Emerging Security Technologies (pp. 56 – 60).

Davis, J. P. Thorniley, S., Gibson, S., & Solomon, C. (2015). Holistic facial composite construction and subsequent lineup identification accuracy: Comparing adults and children. *Journal of Psychology: Interdisciplinary and Applied*, DOI: 10.1080/00223980.2015.1009867

Diamond, R., & Carey, S. (1986). Why faces, are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, 11, 107-117.

Dyfed-Powys Police (2019). Old-fashioned policing and community support solved Pendine murder and manhunt. Press release, 9th April 2019. www.dyfed-powys.police.uk/en/newsroom/press-releases/old-fashioned-policing-and-community-support-solved-pendine-murder-and-manhunt/

Ellis, H. D., Davies, G. M., & Shepherd, J. W. (1978). A critical examination of the Photofit system for recalling faces, *Ergonomics*, 21, 4, 297-307.

Ellis, H. D. & Shepherd, J. W. (1996). Face recall methods and problems. In S.L. Sporer, R.S. Malpass & G. Koehnken (Eds.). *Psychological issues in eyewitness identification*. Hillsdale, NJ: Lawrence Erlbaum. Ch. 5.

Ellis, H. D., & Shepherd, J. W. (1992). Face memory - theory and practice. In M. M. Gruneberg & P. E. Morris (Eds.). *Aspects of Memory*, Vol. 1, The practical aspects (2nd ed.). London: Routledge.

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

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

Fisher, R. P. (1995). Interviewing victims and witnesses of crime. *Psychology, Public Policy, and Law*, *1*, 732-764.

Fisher, R. P., Geiselman, R. E., & Amador. M. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witnesses of crime. *Journal of Applied Psychology*, *74*, 722-727.

Fodarella, C., Brown, C., Lewis, A., & Frowd, C. D. (2015). Cross-age effects on forensic face construction. *Frontiers in Psychology*, DOI: 10.3389/fpsyg.2015.01237

Fodarella, C., Frowd, C. D., Warwick, K., Hepton, G., Stone, K., Date, L., & Heard, P. (2017). Adjusting the focus of attention: Helping witnesses to evolve a more identifiable composite. *Forensic Research & Criminology International*, *5*, DOI: 10.15406/frcij.2017.05.00143.

Fodarella, C., Kuivaniemi-Smith, H. J., Gawrylowicz, J., & Frowd, C. D. (2015). Forensic procedures for facial-composite construction. *Journal of Forensic Practice*, *17*, 259-270.

Fodarella, C., Marsh, J., Chu, S., Wilcock, R., & Frowd, C. D. (2018). Context reinstatement and holistic interviewing techniques to improve the effectiveness of facial composites. *European Association of Psychology and Law (EAPL)*, Turku, Finland, June.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations and interpretation. *Journal of Experimental Psychology: General*, *141*, 2-18.

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

Frowd, C. D., Bruce, V., & Hancock, P. J. B. (2008). Changing the face of criminal identification. *The Psychologist*, *21*, 670-672.

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

Frowd, C. D., Bruce, V., McIntyre, A., Ross, D., Fields, S., Plenderleith, Y., & Hancock, P. J. B. (2006). Implementing holistic dimensions for a facial composite system. *Journal of Multimedia*, *1*, 42-51.

Frowd, C. D., Bruce, V., Ness, H., Bowie, L., Thomson-Bogner, C., Paterson, J., McIntyre, A., & Hancock, P. J. B. (2007). Parallel approaches to composite production. *Ergonomics*, *50*, 562-585.

Frowd, C. D., Bruce, V., Ross, D., McIntyre, A., & Hancock, P. J. B. (2007). An application of caricature: How to improve the recognition of facial composites. *Visual Cognition*, *15*, 1-31.

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

Frowd, C. D., Bruce, V., Storås, K., Spick, P. & Hancock, P. J. B. (2006). *An evaluation of morphed composites constructed in a criminal investigation*. Proceedings of the 16th Conference of the European Association of Psychology and Law, London: IP-PA Publishing, pp. 59-66.

Frowd, C. D., Bunce, L., Madin, C., & Brown, C. D. (2016). Facilitating facial composite construction using a self-administered interview. *Presentation at 9th Annual Conference and Masterclass of the International Investigative Interviewing Research Group (iIRG)*, June, London, UK.

Frowd, C. D., Carson, D., Ness, H., McQuiston, D., Richardson, J., Baldwin, H., & Hancock, P. J. B. (2005a). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal & Criminological Psychology*, *10*, 63-81.

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

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

Frowd, C. D., & Goodfellow, Y. (2017). An investigation into whether poor quality composites are a result of diminished facial recall. In M. Bindemann and A. Megreya (Eds.) *Face Processing: Systems, Disorders and Cultural Differences*. Nova Science.

Frowd, C. D., Hancock, P. J. B., Bruce, V., McIntyre, A., Pitchford, M., Atkins, R., et al. (2010). Giving crime the 'evo!': catching criminals using EvoFIT facial composites. In G. Howells, K. Sirlantzis, A. Stoica, T. Huntsberger and A.T. Arslan (Eds.) *2010 IEEE International Conference on Emerging Security Technologies* (pp. 36-43).

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

Frowd, C. D., Hancock, P. J. B., & Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Psychology (TAP)*, *1*, 1-21.

Frowd, C. D., Jones, S., Fodarella, C., Skelton, F. C., Fields, S., Williams, A., Marsh, J., Thorley, R., Nelson, L., Greenwood, L., Date, L., Kearley, K., McIntyre, A., & Hancock, P. J. B. (2014). Configural and featural information in facial-composite images. *Science & Justice*, *54*, 215-227.

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

Frowd, C. D., Nelson, L., Skelton F. C., Noyce, R., Atkins, R., Heard, P., Morgan, D., Fields, S., Henry, J., McIntyre, A., & Hancock, P. J. B. (2012). Interviewing techniques for Darwinian facial composite systems. *Applied Cognitive Psychology*, *26*, 576-584.

Frowd, C. D., Park, J., McIntyre, A., Bruce, V., Pitchford, M., Fields, S., Kenirons, M., & Hancock, P. J. B. (2008). Effecting an improvement to the fitness function. How to evolve a more identifiable face. In A.

Stoica, T. Arslan, D. Howard, T. Higuchi, and A. El-Rayis (Eds.) 2008 ECSIS Symposium on Bio-inspired, Learning, and Intelligent Systems for Security, (pp. 3-10). NJ: CPS. (Edinburgh).

Frowd, C. D., Pitchford, M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., McIntyre, A., & Hancock, P. J. B. (2010). The psychology of face construction: Giving evolution a helping hand. *Applied Cognitive Psychology, 25*, 195-203.

Frowd, C. D., Pitchford, M., Skelton, F. C., Petkovic, A., Prosser, C., & Coates, B. (2012). Catching even more offenders with EvoFIT facial composites. In A. Stoica, D. Zarzhitsky, G. Howells, C. Frowd, K. McDonald-Maier, A. Erdogan, and T. Arslan (Eds.) *IEEE Proceedings of 2012 Third International Conference on Emerging Security Technologies* (pp. 20 - 26).

Frowd, C. D., Portch, E., Killeen, A., Mullen, L., Martin, A. J., & Hancock, P. J. B. (2019). EvoFIT facial composite images: a detailed assessment of impact on forensic practitioners, police investigators, victims, witnesses, offenders and the media. In A. Stoica, G. Howells, K. McDonald-Maier, A. Erdogan, and T. Arslan (Eds.) *Proceedings of IEEE International Conference on Emerging Security Technologies*, July, University of Essex, DOI: 10.1109/EST.2019.8806211.

Frowd, C. D., Skelton, F. C., Battersby, K., Fodarella, C., & Hancock, P. J. B. (2018). Techniques for substantially improving the effectiveness feature-based facial composites. *European Association of Psychology and Law (EAPL)*, Turku, Finland, June.

Frowd, C. D., Skelton, F. C., Atherton, C., Pitchford, M., Bruce, V., Atkins, R., Gannon, C., Ross, D., Young, F., Nelson, L., Hepton, G., McIntyre, A.H., & Hancock, P. J. B. (2012). Understanding the multi-frame caricature advantage for recognising facial composites. *Visual Cognition, 20*, 1215-1241.

Frowd, C. D., Skelton F. C., Atherton, C., Pitchford, M., Hepton, G., Holden, L., McIntyre, A., & Hancock, P. J. B. (2012). Recovering faces from memory: The distracting influence of external facial features. *Journal of Experimental Psychology: Applied*, *18*, 224-238.

Frowd, C. D., Skelton, F. C., Butt, N., Hassan, A., & Fields, S. (2011). Familiarity effects in the construction of facial-composite images using modern software systems. *Ergonomics*, *54*, 1147-1158.

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

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, *70*, 401-412.

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

Hall, D. F. (1976). Obtaining eyewitness identification in criminal investigations: Applications of social and experimental psychology (Doctoral dissertation, Ohio State University, Columbus). *Dissertation Abstracts International*, *37*, 2569B.

Hasel, L., & Wells, G. L. (2007). Catching the bad guy: Morphing composite faces helps. *Law and Human Behavior, Law & Human Behavior*, *31*, 193-207.

Hill, G. W. (1982). Groups versus individual performance: Are $n + 1$ heads better than 1? *Psychological Bulletin*, *91*, 517-539.

Jenkins, F., & Davies, G. M. (1985). Contamination of facial memory through exposure to misleading composite pictures. *Journal of Applied Psychology*, *70*, 164-176.

Kempen, K., & Tredoux, C. G. (2012). Seeing is believing: The effect of viewing and constructing a composite on identification performance. *Psychological Society of South Africa*, *42*, 434-444.

Koehn, C. E., & Fisher R. P. (1997). Constructing facial composites with the Mac-a-Mug Pro system. *Psychology, Crime & Law*, *3*, 215-224.

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

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

Kuivaniemi-Smith, H. J., Nash, R. A., Brodie, E. R., Mahoney, G., & Rynn, C. (2014). Producing facial composite sketches in remote cognitive interviews: A preliminary investigation. *Psychology, Crime & Law*, *20*, 389-406.

Kuivaniemi-Smith, H. J., Richardson, B., Nash, R., Uther, M., & Frowd, C. D. (2018). The impact of mental reinstatement of context on sketch composites. *European Association of Psychology and Law (EAPL)*, Turku, Finland, June.

Lampinen, J. L., Erickson, B., Frowd, C. D., & Mahoney, G. (2015). Mighty morphin' age progressions: how artist, age range, and morphing influences the similarity of age progressions to target individuals. *Psychology, Crime & Law*, *21*, 952-967.

Lander, K., Portch, E., Frowd, C. D., & Cootes, T. (2017). Can animating face composites make them more recognisable? *European Association of Psychology and Law (EAPL), Belgium*, May 2017.

Laughery, K., & Fowler, R. (1980). Sketch artist and identikit procedures for generating facial images, *Journal of Applied Psychology*, *65*, 307-316.

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

Loftus, E. F. (1997). Creating false memories. *Scientific American*, *277*, 70-75.

Marsh, J. M., Demaine, J., Bell, R., Skelton, F. C., Frowd, C. D., Röer, J., & Buchner, A. (2015). The impact of irrelevant auditory facial descriptions on memory for target faces: Implications for eyewitness memory, *Journal of Forensic Practice*, *17*, 271-280.

Marsh, J. M., Patel, K., Labonte, K., Threadgold, E., Skelton, F., Fodarella, C., Thorley, R., Battersby, K., Frowd, C. D., Ball, L., & Vachon, F. (2016). Chatting in the face of the eyewitness: the impact of extraneous cell-phone conversation on memory for a perpetrator, *Canadian Journal of Experimental Psychology*, ISSN 1196-1961.

Martin, A. J., Hancock, P. J. B., & Frowd, C. D. (2017). Breath, relax and remember: an investigation into how focused breathing can improve identification of EvoFIT facial composites. In G. Howells et al.

(Eds.) *Proceedings of IEEE 2017 Seventh International Conference on Emerging Security Technologies*, 4th – 8th September, University of Kent, Canterbury, UK.

Martin, A. J., Hancock, P. J. B., Frowd, C. D., Heard, P., Gaskin, E., Ford, C., & Hewitt, T. (2018). EvoFIT composite face construction via practitioner interviewing and a witness-administered protocol. In G. Howells et al. (Eds.) *Proceedings of 12th NASA / ESA Conference on Adaptive Hardware and Systems*, 6th - 9th August, University of Edinburgh, Edinburgh, UK.

McIntyre, A., Frowd, C. D., Bruce, V., & Hancock, P. J. B. (2010). Looking at people who are wearing glasses: the impact of glasses on matching photographs of unfamiliar faces and recognizing familiar composite faces. In G. Howells, K. Sirlantzis, A. Stoica, T. Huntsberger and A. T. Arslan (Eds.) 2010 IEEE International Conference on Emerging Security Technologies (pp. 36-43). ISBN 978-0-7695-4175-4.

McIntyre, A., Hancock, P. J. B., Langton, S. R. L., & Frowd, C. D. (2016). Holistic face processing can inhibit recognition of forensic facial composites. *Law and Human Behavior*, 40, 128-135.

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification, *Applied Cognitive Psychology*, 15, 603-616.

Milne, R., & Bull, R. (1999). *Investigative interviewing: Psychology and practice*.

Chichester: John Wiley & Sons, Ltd.

Miner, F. C. (1984). Group versus individual decision making: An investigation of performance measures, decision strategies, and process losses/gains. *Organizational Behavior and Human Performance*, 33, 112-124.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519-533.

Ness, H. (2003). Improving facial composites produced by eyewitnesses. Unpublished Ph.D. thesis, University of Stirling.

Osborn, A. F. (1957). *Applied imagination*. New York: Charles Scribner's Sons.

PACE (2017). Police and Criminal Evidence Act 1984 (PACE). CODE D Revised Code of Practice for the identification of persons by Police Officers. [UK] Home Office.

REF (2014). Providing software, training and support to the police to allow them to identify criminal suspects using facial-composite images. *Research Excellence Framework (REF): Impact Case Study*, <https://impact.ref.ac.uk/casestudies/CaseStudy.aspx?Id=1761>

Richardson, B., Frowd, C. D., Pitchford, M. et al. (in prep). Unattractiveness is good: The advantage of low facial attractiveness for composite production.

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

Shepherd, J. W., Davies, G. M., & Ellis, H. D. (1981). Studies of cue saliency. In G. M. Davies, H. D. Ellis, J. W. Shepherd (Eds), *Perceiving and Remembering Faces*. (pp. 105-132). London: Academic Press.

Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. *European Journal of Cognitive Psychology*, *3*, 137-45.

Skelton, F. C., Frowd, C. D., Hancock, P. J. B., Jones, H. S., Jones, B. C., Fodarella, C., Battersby, K., & Logan, K. (in press). Constructing identifiable composite faces: the importance of cognitive alignment of interview and construction procedure. *Journal of Experimental Psychology: Applied*.

Skelton, F. C., Frowd, C. D., & Speers, K. (2015). The benefit of context for facial-composite construction, *Journal of Forensic Practice*, 17, 281-290.

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

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

Topp-Manriquez, L. D., McQuiston, D., & Malpass, R. S. (2016). Facial composites and the misinformation effect: How composites distort memory. *Legal & Criminological Psychology*, 21, 372-389.

Turner, J., Pike, G., Brace, N., & Vredeveldt, A. (2016). Does making a facial composite contaminate witness memory? Ecological validity and framing an innocent suspect. Presentation at *26th Annual Conference of the European Association of Psychology and Law*, Toulouse, July.

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, 72- 86.

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition, 20*, 291-302.

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

Wells, G. L., Memon, A., & Penrod, S. D. (2007). Eyewitness evidence: Improving its probative value. *Psychological sciences in the public interest, 7*, 45-75.