

## Facial Composites: Forensic Utility and Psychological Research

Graham M. Davies  
School of Psychology  
University of Leicester  
Leicester  
United Kingdom

Tim Valentine  
Department of Psychology  
Goldsmiths' College  
University of London  
United Kingdom

Police composites are impressions of a suspect's facial appearance derived from a witness description. Such disembodied faces stare out from the pages of our newspapers and television screens, coupled with a plea to members of the public to get in touch with the police if they believe they know someone or may have seen an individual bearing a resemblance to the composite. In the United Kingdom, just 10% of composite faces are released to the media. The remainder are used for internal police enquiries: around half are shown to informants familiar with the appearance of local criminals and another third used for house-to-house enquiries in the hope that it will cue a tentative identification (Kitson, Darnbrough & Shields, 1978). Inevitably, many composites end up neglected in police files or thumb tacked to bulletin boards, awaiting the arrest of a suspect by other means. How effective are composite systems in practice? And can they be improved through psychological research?

In this chapter we review four generations of composite systems, together with the psychological research they have provoked. The earliest technique still in use is the artists' impression of a face, rendered from a witness description. The second generation is represented by mechanical systems, such as the Identikit and Photofit, which build up a face from component features (eyes, noses, mouths etc) selected by the witness. A third generation based on software systems, like Mac-a-Mug and E-fit, uses the same principle of witness-guided feature selection, but uses a computer to synthesise and manipulate an image of a face on a video screen. A fourth generation based on the use of genetic algorithms is at the development stage; such systems seek to capitalise on witness's powers to discriminate between whole faces, rather than identify individual features. We conclude by considering whether the fit between the qualities of human memory and the demands of the composite process means that all

systems place an unrealistic burden on the witness: perhaps the quest for the 'perfect' composite system may be illusory.

## ARTISTS' IMPRESSIONS

The use of an artist to sketch a likeness of a suspect from a witness's description has a long history in forensic science. As early as 1911, the technique was used in the hunt for Dr. Crippen, who had fled London, shortly before the remains of his wife were discovered buried in his cellar. The Metropolitan Police circulated an artist's impression of Crippen's current appearance and he was subsequently identified as a passenger travelling under an assumed name on a transatlantic liner. In more recent times, the hunts for the Unabomber and the perpetrators of the Oklahoma and Bali bombings have also involved widespread publicity for artists' impressions (Taylor, 2001).

### Construction Methods

Despite the publicity surrounding their work, there is little consensus among police artists about the appropriate method for constructing a likeness and no international standards for such sketches. The International Association for Identification has a Forensic Art Certification Board and the American FBI runs an annual training course, but the influence of such bodies appears limited (Domingo, 1984).

Most artists work directly with the witness, but FBI operatives are taught to work at a distance from a description provided by a field officer (Clifford & Davies, 1989). A number of experienced artists have written of their own methods (e.g. Cormack, 1979; Homa, 1983; Taylor, 2001), but their views differ on such matters as whether photographic reference material should be used or whether the artist should rely upon freehand drawings; whether caricature should be used to emphasise distinctive features, and the time to be allocated to capturing characteristic expressions (Davies, 1986b).

Taylor (2001) has described in detail her own approach to obtaining a likeness. In the Pre-Interview Stage, the artist and the investigator review the circumstances of the crime and the opportunities the witness had to view the suspect. Drawings should not be attempted if the witness had very limited or fragmentary views. In the Rapport Building Stage, the artist gets to know the witness as a person and explains the goal of composite art. The artist is aiming for an impression not a finished portrait. In the Initial Drawing Stage, the artist elicits a detailed verbal description, which forms the basis of an outline drawing, with priority given to features emphasised by the witness. At the Fine-Tuning Drawing Stage, the drawing is progressively refined; reference material in the form of mugshots exemplifying particular features or groups of features may be shown to help the witness. The final Finishing Touches involve a review of individual features and perhaps attention to expression. The witness may be encouraged to give a score out of 10 for degree of likeness. According to Taylor, police artists take from one to three hours to evolve a satisfactory drawing.

### Research on the Effectiveness of Artist's Sketches

Apart from demonstrations of the effectiveness of caricature (Benson & Perrett, 1991; Rhodes, 1996), little empirical research appears to have been conducted

on the assumptions and recommendations of individual artists (but see Davies, 1986b; Davies & Little, 1990). Anecdotal accounts testify to the success of individual artists in capturing likeness (Garcia & Pyke, 1977; Boylan, 2000), but there appear to have been no systematic attempts to gauge their overall effectiveness under police operational conditions. It would be difficult to arrive at an overall estimate, given the widespread differences in the way that individual artists work. To be effective, a sketch artist must not only be good at portraiture but also possess the necessary interviewing skills to elicit relevant information from the witness (Taylor, 2001). Some artists regularly employ the Cognitive Interview to elicit the necessary facial description (Frowd, Carson, Ness, McQuiston-Surrett, Richardson, Baldwin & Hancock, in press). Not surprisingly, a combination of interviewing and artistic talents is rare and such individuals tend to be brought in by the police on an ad hoc basis for high profile cases. The United States has over 500 Sheriff's Departments, but only 18 full-time artists (Poole, 2004). According to one US sheriff, 'It is a dying art' (Penserga, 2003) and for most cases, police increasingly rely upon mechanical or computer-based composite production systems.

## MECHANICAL SYSTEMS

### The Identikit

The need for a uniform system that could reproduce facial resemblance without the intervention of a skilled police artist was recognised by Hugh MacDonald, a Californian police officer, who introduced a device called the Identikit in 1959. The original Identikit consisted of some 568 drawings of different facial features: chins; eyebrows; eyes; hairstyles; lips, and noses reproduced on transparent acetate sheets. MacDonald advocated that witnesses be asked to provide a verbal description of each feature in turn. The operator would then select the acetate foil that best fitted the description and the foils superimposed to yield a composite face. The witness could then refine this first composite by exchanging and adjusting features until a satisfactory likeness emerged. Foils were number coded, enabling the rapid transmission of likeness information from one force to another in the days before facsimile transmission. No systematic investigation seems to have been undertaken of the level of accuracy achievable by the system or of its operational effectiveness, although there are striking stories of isolated successes, both in the United States (Sondern, 1964) and the United Kingdom (Jackson, 1967).

### Photofit

One perceived weakness of the original Identikit was the absence of realism in the monochrome drawings. Subsequent research has shown that the naming of even famous faces from simple line drawings is very poor. It is necessary to add the depth cues and shading normally present in photographs before such drawings are readily identified (Davies, Ellis & Shepherd, 1978b; Bruce, Hanna, Dench, Healey & Burton, 1992).

In 1970, the British inventor Jacques Penry persuaded the police in the United Kingdom to adopt and develop a composite system based on actual photographs of facial features: the Photofit system. In its final form, Photofit, like Identikit, contained examples of some 560 facial features: hairstyles; pairs of eyes and eyebrows; noses; mouths and chins of which hair formed the single largest group (213 different styles). Each example was printed on thin card and could be superimposed, jigsaw fashion, within a special frame to produce a composite face. Complimentary to the features

was a directory or 'Visual Index' reproducing each of the features in miniature for consultation by the witness. Like the original Identikit, Photofit also contained a range of accessories such as hats and spectacles to enhance the final likeness.

Photofit was supplied with no specific instructions as to use, apart from a book illustrating Penry's approach to physiognomy (Penry, 1971). However, most operators were taught to begin by eliciting a verbal description from the witness whose attention would then be directed by the operator to particular features in the Visual Index that appeared to correspond to the description. The selected features would then be assembled in the frame and the initial likeness shown to the witness for comment and subsequent amendment. Plain acetate sheets and wax pencils were also provided for amending the image through the addition of scars, tattoos etc. Like the Identikit, there were no formal trials of the system, though its introduction was overseen by a working party of police identification personnel (King, 1971).

### System Development

After its introduction, Photofit spread to some 20 countries and Identikit was also extensively marketed, latterly in a revised form that featured photographic levels of realism in its features (Identikit II; see Owens, 1970). Additional kits were produced for rendering likenesses of women as well as men and to model different-race faces, such as Asian and African-Caribbean. Police forces in other countries developed their own systems, such as France ('Portrait Robot'), Germany and Italy, but all were based on the same principal of the recognition of individual features and their fusion into a composite face (see Allison, 1973; Davies, 1981 for reviews).

### Early Evaluations of Photofit

An initial attempt to gauge the likely accuracy of the Photofit kit was reported by Ellis, Davies and Shepherd (1975). In one study, witnesses worked with a trained operator to reproduce a likeness after briefly viewing a photograph of one of a number of white male targets. The resulting composite was then viewed by panels of judges who attempted to choose the correct face from an array of 36 different faces. The accuracy of the judges at this task was generally poor: while there were isolated examples of likenesses which were readily recognised, overall accuracy was generally poor, with just 12.5% of judges' first selections being correct, which increased to 25% if their second and third choices were taken into account.

Davies, Ellis and Shepherd (1978a) asked participants to make Photofit composites of two faces, one immediately following observation and a second after a delay of one week. Degree of likeness of the composites was assessed both by rating scales and an identification task. Overall level of accuracy was again poor and there was no measurable change in quality of likeness between composites made immediately and those made after a delay, despite a follow-up study confirming that recognition memory for the faces had significantly deteriorated in the interval. The authors concluded that this was further evidence for the insensitivity of the system.

Ellis, Davies and Shepherd (1978a) compared Photofit composites made in the presence of a photo of the target face with those made from memory. Again, no differences in rated quality of likeness emerged as a result of viewing condition, a finding again suggestive of low sensitivity in the system. In an attempt to probe memory for the face independent of the composite, the witnesses themselves made sketches of the faces. These drawings showed significant differences in rated quality between those made from memory and those made in the presence of the target, again suggesting gross insensitivity in the composite system.

Two exceptions to this insensitivity rule concern the impact of race and age. Facial recognition within racial groups is generally better than across groups (e.g. Chance & Goldstein, 1996; Chiroro & Valentine, 1995). Ellis, Davies and McMurrin (1979) reported that composites of a white face made from memory by black South African participants were matched by judges to the correct faces significantly less accurately than those made by white Scots. However, there was no corresponding advantage for black witnesses on the black faces: both groups produced composites that were poorly matched against the correct faces by the judges. The authors attributed this to the smaller range of features included in the black Photofit kit. However, all the judges were white and the possibility that a black panel might have produced a different pattern of results cannot be excluded.

Children show marked developmental improvements in their ability to recognise faces with age (see Davies, 1996 for a review). Flin, Markham and Davies (1989) asked children to briefly observe a photograph of a male face before compiling a Photofit from memory. Both the initial verbal descriptions and the subsequent composites produced by children aged 8-9 years were matched to the correct photographs significantly less accurately by adult judges than those made by 11-12 year olds. The accuracy of the verbal descriptions produced by children of different ages was not significantly linked to the quality of the composites they produced, suggesting that verbal description and composite production may draw upon rather different skills.

Mention of the preliminary verbal descriptions raises one of the most surprising results for Photofit to be reported by the research team. One of the assumptions of all composite systems is that the visual image of the face is a more powerful aid to identification than the verbal description from which it is derived: an impression of a face should be worth a thousand words. Christie and Ellis (1981) compared the relative effectiveness of the initial verbal description elicited from experimental participants with the finished Photofit composite as a guide to likeness. Verbal descriptions were a consistently better guide to likeness than the Photofit composites. Moreover, a combination of description plus composite was no better than description alone.

Taken together, the results of these experimental studies suggest that Photofit is a very imprecise tool for conveying facial likeness. Is this result typical of all mechanical composite systems or confined solely to Photofit?

### Evaluation of the Identikit

The only other system to be extensively researched was the original Identikit, studied by Laughery and his colleagues. Laughery and Fowler (1980) had volunteers converse with a target for 7-8 minutes before working with a trained Identikit technician or a police artist to produce a likeness of the target's face. Subsequently, technician and artist constructed a likeness with the target present. The composites were then assessed for degree of likeness by rating scales and a computerised search task of a database that included the target faces. Irrespective of the race or gender of the witness, the ratings showed a very similar pattern. Artist's sketches were judged as superior to Identikit composites. Moreover, while sketches made from memory received lower ratings than those made in the presence of the target, no such difference was found for the Identikit, precisely paralleling the findings on Photofit. Both artists and the Identikit performed poorly in the computerised search task. Identikits were at chance except for a subgroup of composites that received

particularly high ratings of likeness, but even here, high rated sketches were superior (Laughery & Smith, 1978).

Other results also show parallels to those reported for Photofit. For instance, for delay, McNeil, Wray, Hibler, Foster, Rhyne, and Thibault (1978) could detect no change in quality for Identikits made after three weeks compared to those constructed immediately after observation (though a later study by Green and Geiselman, 1989, did detect a decline in quality with a delay after a week with Identikit II). Like Photofit, the Identikit did show sensitivity to age. Shwartz-Kenney, Norton, Chalkley and Jewett (1996) had children aged 5-6 or 8-9 years of age interact with a stranger for 15 minutes before attempting to build a likeness of his face. Identikit portraits made by the older children were rated as better likenesses compared to those of the younger children, with no effect for gender of child.

### Possible Limitations on Experimental Studies

From these experimental studies, it appears that mechanical composite systems are of questionable forensic value. However, before condemning such systems wholesale, some of the limitations of the experimental work need to be underlined.

For instance, many of the Photofit studies used very brief exposure intervals and photographs rather than an actual person as targets. It could be argued that composites are rarely compiled after such a brief exposure to the suspect. However, extending the exposure interval led to no demonstrable increase in the quality of likeness of the composite (Ellis et al., 1978). Equally, while the use of an actual person as a target provides the witness with greater depth and shape cues than a photograph (Bruce, 1988), Ellis, Davies and Shepherd (1976) could detect no difference in composite quality when live and photographic targets were directly compared. Further, all of Laughery and Fowler's studies of the Identikit used long exposure intervals combined with actual persons as targets and results were just as disappointing as for Photofit. Another criticism is that in most of the studies reported, accuracy was assessed by such methods as ranking composites in terms of degree of likeness, sorting or matching composites against photographs of the target faces. These methods certainly lack the forensic realism of the identification from an array, the task employed by Ellis et al (1975), but they produce accuracy scores that are reliable and are significantly intercorrelated, suggesting that they are tapping a common underlying process (Davies et al, 1978a). Finally, most of the studies cited make no attempt at forensic realism: the witnesses do not believe a crime is taking place and there is little personal investment in constructing an accurate likeness of the 'offender'. There is certainly room for more ambitious experimental attempts at simulated crimes, though evidence from the field studies reviewed below does not suggest accuracy of witnesses is likely to be enhanced by real crime settings.

A more subtle point concerns the choice of dependant variables. Operationally, police do not necessarily seek a pinpoint likeness, but rather to isolate a subset of persons from whom the suspect is drawn and equally importantly, to eliminate suspects who bear no resemblance to the suspect. Thus, if a witness compiles a round, podgy-faced Photofit, investigators may switch enquiries away from lean-faced suspects to focus on the fuller faced (Davies, Ellis & Shepherd, 1985). How effective are such composite systems at conveying such type-likeness information? Christie, Davies, Shepherd and Ellis (1981) explored this issue by asking subjects to attempt to match Photofit composites gathered from memory under experimental conditions to an array of photographs of men's faces, one of which was

always the target. The faces had previously been assessed for degree of likeness to each other, by asking other judges to sort the faces into groups on the basis of likeness and then using hierarchical clustering analysis to isolate groups of physiognomically similar faces. When the matching scores were assessed by the traditional criterion of perfect likeness, only 23 % of choices proved correct. However, when the criterion was relaxed to include a correct type likeness, then some 48 % were satisfactory. Clearly, there is information present in the average composite which can be forensically useful, but the 52 % of composites that failed to meet even the type likeness criterion must continue to be a source of concern as such composites could lead police to disregard the actual perpetrator.

One final consideration concerns the skills of the operator. A composite system is only as good as the technician using it. As has been noted, Photofit contained no explicit instructions on how it was to be deployed operationally and training courses for operatives, with input from psychologists, were a comparatively late development (Davies, Shepherd, Shepherd, Flin and Ellis, 1986). Evidence for the value of expertise in compiling composites emerges from later studies that compared the quality of composites made in the presence of a photograph of the target and those made from memory. Early studies of both the Identikit and Photofit suggested no difference in assessments of quality underlining the apparent insensitivity of the systems. However, later studies of Photofit using a very experienced operator, who had compiled many hundreds of composites, produced reliable differences in quality between composites made from memory and those from view (Christie et al, 1981). The same expert operator took part in a further study when her skills were assessed against a novice operator, familiar with the mechanics of the kit, but with little practical experience of its use (Davies, Milne & Shepherd, 1983). Both were required to compile two target faces described to them by individual witness subjects. The composites produced by the expert were rated as better likenesses and were sorted more accurately than those made by the novice. Analysis of the process of composite production suggested that the expert took longer over the verbal description phase and tended to elicit richer and more elaborate descriptions compared to the novice. This strategy had also been noted by Laughery, Duval and Wogalter (1986) among successful police artists.

Will real witnesses to crime do any better than research volunteers in the laboratory? The most systematic survey on the operational effectiveness of Photofit was conducted by the British Home Office (Kitson et al., 1978) and suggests that the laboratory findings are broadly representative of field outcomes. Over a sixth month period, Kitson et al. followed up some 729 composites made in the course of police enquiries by 15 different police forces. After two months, 140 cases had been cleared up and the investigating officer was contacted to establish what role Photofit had played in this. According to the officers, in some 5% of cases Photofit was entirely responsible for solving the case: the image produced by the witness was immediately identified and the suspect arrested. In 50% of cases, it was 'very useful' (17%) or 'useful' (33%) in solving the crime: typically a good type likeness that narrowed the focus of the enquiry. However, in 45% of cases, the composites proved either 'not very useful' (20%) or 'no use at all' (25%). These would be examples of composites that diverted enquiries and wasted police time. A later survey of Photofits produced by the Metropolitan Police produced rather similar proportions, albeit from a much smaller sample of resolved cases (Bennett, 1986). Research suggests that these disappointing findings are not unique to Photofit. Levi (1997) reported that of 243 cases in which Identikit II was used by the Israeli police, 54 led to convictions, but

only 5 were deemed to have been significantly aided by the presence of the composite. Experimental evidence suggested that the 'successful' composites were no better guides to likeness than those which did not lead to convictions.

### Evaluation of the Mechanical Systems

From these findings, it is hard to argue that the laboratory research paints an over-pessimistic picture of the forensic utility of mechanical composite systems. Publicised successes need to be balanced against complete failures to render an effective likeness. The particular combinations of witness characteristics, suspect appearance and viewing conditions likely to lead to a good quality composite remain elusive. In a study where witnesses made pairs of Photofit composites, the rated quality of one likeness was essentially unrelated to the other (Davies et al 1978a). It remains to be asked why mechanical systems are so relatively poor at rendering likeness.

One problem is the range and representativeness of the features in the kits. While the number of features appeared large and the possible different combinations impressive, the features represented reflected intuition rather than the result of any systematic research. It was evident that while the kit could make some faces well, others were impossible with the supplied set of parts (Ellis et al., 1976). Research employing multi-dimensional scaling of the likeness judgements made on large populations of faces suggested that age, face shape and quality and distribution of hair are important dimensions of judgment of likeness (Ellis, 1986; Shepherd, Davies & Ellis, 1981). As Bruce and Young (1998) have observed, age and face shape are global dimensions involving multiple features that are very difficult for mechanical systems to model. One common complaint of Photofit operators was the lack of youthful features in the kit, which gave most composites a middle-aged look (Davies et al., 1985).

One answer to this was the introduction of the Aberdeen Supplement to the Photofit male kit, which included an additional 80 features selected from the female kit and judged as sufficiently androgynous to pass as 'young' masculine features. Despite the disproportionate number of hair sections included, shifting fashions in hair styles has always presented a particular difficulty for composite systems. The Aberdeen Supplement included a number of female hairstyles to try to cope with the vogue for longer hair among younger men in the '80s. However, these were stopgap measures, which did not address the wider issues of achieving global change in faces created by all mechanical composite systems.

Another difficulty inherent in mechanical systems was the way in which the use of fixed components inevitably constrained what aspects of the face could be changed. Thus, the distance between the eyes or the eyebrow to hairline distance can have a major impact on degree of likeness (Haig, 1986). However, mechanical systems like Photofit and Identikit cannot readily accommodate changes of this kind. In Photofit, eyes and eyebrows came as a single piece and it was up to the operator to try to amend the composite with a wax pencil if a witness liked the eyes, but took exception to the brows or vice versa. Global changes, such as making a face longer or wider, involved either laborious exchanges of individual features or very extensive overdrawing on top of the basic composite which were not always successful in achieving the appropriate outcome (Gibling & Bennett, 1994).

Finally, there was the rationale of the systems, which assumed that witnesses could readily parse a remembered face into component features and relate such features to the foils in the Identikit or the examples included in the visual index of



Photofit. Research on the process of face recognition suggests that faces are normally encoded not as a string of features, but rather, as an overall gestalt in which feature information is subsumed within a general impression of the face as a whole (Tanaka & Farah, 2003; Rakover, 2002). Encoding a face in terms of an overall impression (configural processing) is an ideal strategy for facial recognition, but may hinder the recall of individual features where a feature-based approach is required (Wells & Hryciw, 1984). A demonstration of the difficulties of extracting feature information accurately from memory of an overall face was provided by Davies and Christie (1982). Participants had an extended opportunity to observe a male target before rating the similarity of 30 mouths drawn from the Photofit kit. Judgements were made from memory, viewing the mouths as isolated features or embedded in a composite face resembling the target. Ratings in these two conditions were essentially uncorrelated. However, if judges then made ratings on the features in the presence of the target face, these ratings were highly correlated with those made when the mouths were placed in a composite face, but not with the features in isolation. This result implies that judgements of features from memory are more veridical when made within a schematic face than in isolation.

It appears that the very process embodied in mechanical systems of synthesising a completed face from judgements on individual features may be psychologically flawed. The face is more than the sum of its parts and to achieve maximal likeness, witnesses need to be able to manipulate a total face rather than make discriminations based on isolated feature information. The ability to make such global changes and to store large and more representative repertoires of features required the abandonment of mechanical methods for the versatility and power of the modern computer.

## SOFTWARE SYSTEMS

Gillenson and Chandrasekaran (1975) had demonstrated the potential of computer graphics to provide a composite tool of great versatility. The Computer-Aided Design Centre (CADC) in Cambridge built a working prototype system using a powerful mainframe computer at the request of the British Home Office in 1978. The system used the features from the Photofit system in digitised form that could be called up onto a screen. Programs to warp or stretch features or groups of features provided additional flexibility and an averaging algorithm eliminated the skin tone boundaries between components to produce a more life-like 'face' (Kitson et al., 1978). However, results from early trials that compared degree of likeness achieved relative to a conventional Photofit kit were disappointing: composites produced from memory with the CADC prototype were no more accurately recognised than those made by the traditional mechanical method (Christie et al, 1981) and further progress had to await the arrival of the desktop computer and cheaper, more versatile graphics packages.

A number of manufacturers entered the market with rival composite systems (see Clifford & Davies, 1989; Shepherd & Ellis, 1996 for reviews). Two representative systems, which have been subject to extensive research, are Mac-A-Mug Pro, designed for the Apple Macintosh computer and the E-fit system that utilises the Windows technology of the PC. Both are based on the traditional approach of synthesising the desired face from a library of features.

### Mac-a-Mug Pro

Mac-a-Mug Pro (Shaharazam, 1986) uses a modest database of line-drawn facial features (184 hairlines; 117 eyebrows; 13 ears; 65 noses; 80 mouths and 45 chins). However, much greater variety is claimed through the use of specialised editing processes. Features, for instance, can be enlarged or shrunk, age lines and skin complexion darkened, eyes moved further apart, hairlines and facial hair trimmed or extended. The manufacturers offer no guidance as to how best the system should be employed, but most technicians begin by eliciting a brief verbal description, which is then used as a guide to relevant features that may be viewed on screen or in a visual reference catalogue. Once features have been selected and modified if necessary, a composite face is synthesized on screen for the witness's evaluation; further fine-grain changes can be accomplished, using specialised graphics packages (Koehn & Fisher, 1997).

Cutler, Stocklein and Penrod (1988) compared the value of photographs of targets and Mac-a-Mug composites as aids to identifying faces in a photographic array. An experienced operator who had continuing reference to photographs of the targets compiled the composites. Participants searched for the targets in the presence of the likenesses or from memory. Judgements were well above chance in all conditions, with those made in the presence of the likenesses being superior to those made from memory, but the composites were as effective as the photographs in the memory condition. This study demonstrates that under ideal circumstances, the Mac-a-Mug system is capable of generating a highly recognisable composite. Wogalter and Marwitz (1991) used volunteer witnesses to compile six composites of different target faces, first from memory and later from a photograph. Composites made from a photograph were rated as better likenesses than those made from memory, suggesting a basic sensitivity in the system, though this result was not repeated when judges attempted to match targets to sample faces. In a study of greater forensic realism, Koehn and Fisher (1997) allowed participants to meet a stranger before being asked to compile the stranger's face using Mac-a-Mug Pro. The resulting composites were then rated for degree of likeness: 69% of the composites shared the lowest two ratings on a 10-point scale. When judges attempted to use the composites to match to the target face in a six-photo array, just 4% were correctly matched. When other judges performed the same task using composites of the target generated by the trained operator from life, the matching score rose to 77%, emphasising that the problem with reconstruction did not lie in the inability of the system to make the requisite face, but in witness' memory. Contrary to earlier findings reported by Davies and Milne (1985) for Photofit, instructions designed to encourage visualisation and context reinstatement were no more effective than standard instructions.

Similar disappointing results emerged from a series of experiments reported by Kovera, Penrod, Pappas and Thill (1997). An important feature of their studies was the use of familiar faces as targets, rather than total strangers. Students compiled composites of former teachers and classmates. These were then shown to fellow students, familiar with the targets, who attempted to discriminate them from unfamiliar composites. Judgements were made in terms of familiarity, confidence and where possible, naming. Despite being informed of the origins of the composites, just 3 out of 167 names offered by judges were correct! Moreover, constructor's ratings of familiarity of the target and quality of the composite were unrelated to identification accuracy on any measure. The authors concluded 'In the light of the results from this study, it appears that the Mac-a-Mug system's facility for producing

recognisable composites under laboratory conditions is severely limited.’ (Kovera et al., 1997, p.241).

### E-Fit

Are the negative results unique to Mac-a-Mug Pro or are they common to all face construction software? Both Koehn and Fisher (1997) and Kovera et al. (1997) speculate that a composite system which made more concessions to a configural rather than a feature-based approach to face construction might fare better when witnesses must construct faces from memory. One system that explicitly seeks to accommodate a configural approach is the E-fit system (Aspley Limited, 1993), used extensively in the United Kingdom and elsewhere. E-fit owes much to the CADC system and unlike Mac-a-Mug, uses features of photographic quality. It is also marketed with explicit guidance on its use and regular training courses are offered (Clark, 2000). The method recommended involves an extensive initial interview to establish whether the witness saw enough of the suspects face to make an attempt at a composite worthwhile, which may involve the use of the Cognitive Interview to facilitate witness recall (Finger & Pezdek, 1999). Then witnesses provide a verbal description of the suspect’s facial features, cued by on-screen multiple-choice questions. These answers in turn drive an algorithm that selects the most appropriate features from the E-fit database and these features are displayed as a total face. The witness can then amend this by scrolling through alternative features within the context of the face until an acceptable likeness emerges. Finally, fine grain changes, such as trimming or lengthening hair or the addition of scars or tattoos can be accomplished, using a standard graphics package.

Davies, van der Willick and Morrison (2000) compared the effectiveness of E-fit to the old Photofit system in constructing familiar and unfamiliar faces. The composites were then shown to a panel of judges familiar with the appearance of the targets, who rated them for familiarity, providing names where possible and finally, attempted to match the composites to photographs of the targets. Performance across all three tasks produced a similar pattern. Consistent with earlier findings from Mac-a-Mug, familiar faces constructed in E-fit in the presence of the target were disproportionately better than any other condition. Judges averaged 83% accuracy for matching such composites to correct targets. However, in the memory conditions, whether composites were of familiar or unfamiliar faces, no discernable difference in performance between E-fit and Photofit was detectable. As in the Kovera et al (1977) study, naming was problematic. None of the composites made in either system by witnesses initially unfamiliar with the appearance of the target were ever named correctly, though judges gave many incorrect identifications.

Findings interpreted as more favourable to E-fit were reported by Brace, Pike and Kemp (2000). An experienced E-fit operator constructed pairs of composites for a series of 48 famous personalities, the first from memory and second with the aid of a reference photograph. A second condition involved a witness describing the same faces to the operator, first from memory and then with the photograph present. Judges were able to correctly identify 35% of the pairs of composites made by the operator and 25% of the pairs made from witness descriptions. However, the design precluded judgements being provided exclusively on composites made from memory and rates of incorrect identifications were also not reported. When given feedback as to the identity of the person described, judges rated composites made by the witness from memory as poorer likenesses than those made with the aid of a photograph. Less favourable findings were reported by Davies and Oldman (1999). Witnesses assisted

an operator in constructing one of four famous faces, first from memory and then with a reference picture continuously present. As in the Brace et al study, E-fits made from memory received lower rankings than those made from view. However, when judges were asked to name the persons, just 10% of the composites made from view were identified and less than 6 % of those made from memory. Moreover, this was coupled with a 25.2% false naming rate.

One way of boosting identification rates might be to publish all witnesses' attempts at a likeness, either as a set or in the form of a single image, morphed from the constituent likenesses. However, placing a good likeness with three poor ones reduces identification rate compared to one good likeness alone (Brace, Pike, Kemp, Turner & Bennett, 2001). Morphed composites appear to have advantages over a single good likeness when recognising familiar faces, but this is lost for unfamiliar faces made from memory, arguably the most forensically relevant condition (Bruce, Ness, Hancock, Newman & Rarity, 2002).

Could a changed method of composite construction more successfully foster retrieval of configural information? Certainly, there was no evidence in the Davies et al (2000) study to suggest that the approach encouraged by E-fit was any different to the traditional Photofit; they were indistinguishable in terms of the order of construction and the time taken to select features.

### Evaluation of Software Systems

Software-based facial reconstruction systems allow much greater control over the manipulation of the configural properties of a face than was possible with mechanical systems. Credible and readily identifiable composites can be built by these systems, provided a reference photograph of the target is available to the operator or witness at the time of construction. Problems over the range and representativeness of features seem to have been solved, at least for white Caucasian male faces. However, the problems of constructing a good likeness from memory appear to remain for most witnesses. In the Davies et al (2000) study, facial composites produced from memory by a sophisticated software system were of no greater utility than were composites produced by an old mechanical system.

Why do such software systems produce such disappointing results under laboratory test? One weakness could be the continuing reliance on a logical rather than a psychological analysis of face encoding (Davies et al., 1985). A more successful approach might start from a thorough analysis of how faces are perceived and remembered and then use these insights to construct a system. This is the premise of the fourth generation of composites, which attempt to evolve a remembered facial image within a face space.

## THE FOURTH GENERATION: EVOLVING FACES.

The task of building a facial composite requires that the witness synthesise a given face by retrieving individual facial features. However, as has been noted, the available evidence suggests that face perception does not normally involve analysing the face into its constituent parts. The conflict between the nature of facial encoding and task demands may be the underlying cause of the poor utility of mechanical and software systems.

### Face-space

A face similarity space, commonly referred to as 'face-space', provides a useful framework for understanding face recognition. The central idea is that faces are

encoded in a multidimensional similarity space (Valentine, 1991a, b; 1995; 2001). This framework enables face-processing phenomena to be understood in terms of the similarity within a population of faces, without necessarily defining the dimensions on which faces are encoded. Face-space has provided a useful, single framework for understanding disparate face processing phenomena, including the effects of distinctiveness and race (Byatt & Rhodes, 1998; Chiroro & Valentine, 1995; Valentine & Endo, 1992), inversion (Valentine, 1991), caricature (Lee, Byatt & Rhodes, 2000) and the development of face recognition (de Haan, Humphreys & Johnson, 2002). Two recent theoretical developments have now been applied to develop a fourth generation of facial composite systems. First, principal component analysis has been used to implement a face-space and, second, genetic algorithms have been used to search the space to converge on a desired facial likeness.

#### Use of principal component analysis to implement face-space

Principal component analysis (PCA) can be used to extract a set of dimensions (known as ‘eigenfaces’) from a sample of faces on which they can be encoded (Sirovich & Kirby, 1987; Turk & Pentland, 1991). The eigenfaces can be used to encode and re-construct the appearance of the original sample and new faces from the same population. In effect, the principal components provide the dimensions of the face-space. More precisely, this similarity space is an image-space as the principal components are derived from one specific image of each face. Each eigenface is ‘holistic’ because it codes variance across the entire image; faces are not encoded in terms of their parts. Some principal components can be interpreted, for example appearing to code gender (O’Toole, Abdi, Deffenbacher & Valentin, 1995) but many components are not interpretable. The eigenface representation shows an important property postulated by the face-space framework: faces closer together in the PCA space are perceived as more similar to each other (Tredoux, 2002).

A face can be reconstructed by combining the eigenfaces (or principal components) in the correct proportions. Any face, from the same population as the sample used to derive the PCA, can be coded as a set of weights of a given set of eigenfaces. Thus artificial faces can be constructed by any novel combination of weights.

There are some caveats that should be added. First, faces can be viewed as having two aspects to their appearance: texture and shape. Texture is given by the greyscale or colour information in the image of a face. Shape is defined by the position of landmark features (e.g. the corners of eyes and mouth). The construction of synthetic faces from PCA only works well if the faces in the sample are ‘shape-free’; that is the landmarks are located at the same position in each face image. Therefore, all of the fourth generation composite systems morph faces to the average shape of the faces in the sample using a technique introduced by Craw and Cameron (1991). PCA is carried out separately on the texture and shape information. Shape and texture can be combined using a further PCA into an active appearance model which gives a single set of optimally compact parameters (Cootes, Edwards & Taylor; 1998; Cootes & Taylor, 2001).

A second caveat is that PCA does not re-construct the texture of hair accurately. The solution adopted in both Evo-fit and Eigen-fit involves selecting a hairstyle from a database in the same manner of earlier face reconstruction systems, prior to commencing the evolutionary search, and restricting the PCA to the face excluding the hair. Fortunately, the style, length, texture and colour of hair are attributes that witnesses find relatively easily to describe verbally.

### Evolving faces to navigate the face-space

PCA can be combined with a genetic algorithm (GA) to converge on the desired facial image. The genetic algorithm is so-named because it uses two principles of evolution: random variation (or mutation) and selection. The construction of a facial composite begins by the generation of a random set of (artificial) facial images within the PCA space. The witness then selects the image or images that are most similar to the appearance of the culprit. In the initial set there will be a wide range of facial appearances and none are likely to closely resemble the culprit. The selection made by the witness is then used to 'breed' a new set of images introducing mutations around the 'parent' face or faces. The process is repeated iteratively, with each successive 'generation' becoming more similar to the culprit and to each other. The process continues until the witness cannot choose because all of the faces resemble the culprit equally well, or it becomes clear that the GA has failed to converge on the desired appearance.

### Systems under development.

Three research teams are developing facial reconstruction systems based on these principles. Hancock, Frowd and colleagues (Stirling University, Scotland) are developing a system called 'Evo-fit' (Hancock, 2000). Solomon and colleagues (University of Kent, England) are developing a system known as 'Eigen-fit' (Gibson, Pallares Bejarano & Solomon, 2003). Tredoux, Rosenthal and colleagues (University of Cape Town, South Africa) are developing a system known as 'ID' (previously 'E-face'; Tredoux, Rosenthal, Nunez & da Costa, 1999). Both Tredoux and Solomon recombine shape and texture into an active appearance model, allowing the witness to choose between facial images that differ in shape and texture. Hancock uses separate PCA spaces of shape and texture. Witnesses are asked to choose a best-likeness from both a set of images that vary in shape and another set of images that differ in texture. It is possible to select the texture of one face with the shape of another.

The challenge is to develop a system that produces life-like images, converges quickly on the desired appearance and is easy for the witness to use. Quick convergence and ease of use can be conflicting requirements. The witness may provide rich information, for example by providing a numerical rating of every image in a 'generation' for similarity to the target. However the demands placed on the witness are relatively high. Alternatively, the witness may be asked simply to pick the face in a set that is most similar to the target appearance. This task is easier for the witness, but provides less information to guide the evolution of the next generation and may require many generations to produce a recognisable reconstruction. Evolution can arise from crossover (e.g. between the appearance of two 'parents') and mutation (random variation of single appearance from one generation to the next). Algorithms that allow crossover and mutation will tend to produce more variation within each generation.

Gibson, Pallares Bejarano and Solomon (2003) identify three evolutionary algorithms:

Scale Rating (SR) – All of the images in each generation are rating on a numeric scale for similarity to the target. Two faces are selected to breed the next generation enabling both crossover and mutation. Hancock (2000) used a similar approach.

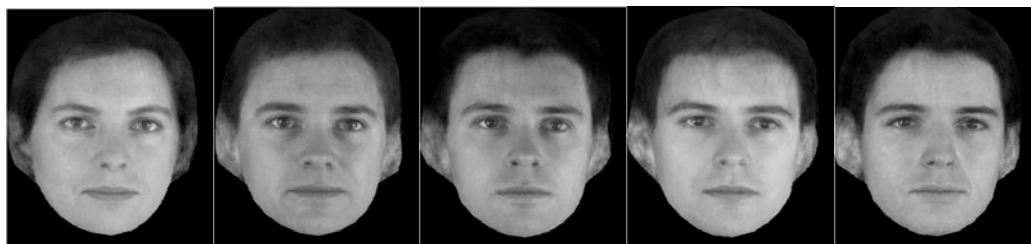
Select Multiple Mutate (SMM). The witness chooses the best –likeness. This image is then reproduced with random mutation in all but one of the faces of the next generation. Tredoux *et al.* (1999) describe a similar approach that they term Population Based Incremental Learning (PBIL).

Follow-The-Leader. One new face is displayed with the current best-likeness. The witness simply chooses the best-likeness of the two faces. The new face displayed at each iteration is produced by breeding the current best-likeness with a new face. The recent evolutionary history is used to determine the future trajectory of the evolution. If the process has followed a well-defined direction, a preference for this direction can be used in subsequent generations.

The allure of using genetic algorithms lies in the gradual holistic changes to faces that exploits the witness’s natural ability to recognise the culprit’s face, rather than requiring the witness to undertake the very difficult task of verbally describing facial features. However, sometimes a witness will comment that the likeness would be improved by a change to a specific feature (e.g. a smaller chin, thicker eyebrows). The evolutionary nature of a genetic algorithm makes it impossible to make a specific change to a local feature easily. Therefore all of the systems described include a facility to make specified changes to the features or position of features of the current best likeness. The modified face can then be used to breed a new generation.

#### Evaluation of GA systems.

All of the fourth generation systems are still under development so there are few evaluations of their performance to date. Gibson *et al.* (2003) report trials based on simulated witness behaviour in which the Select Multiple Mutate algorithm required 150 iterations, and the Follow-The-Leader algorithm required 350 iterations to produce a ‘quasi-perfect’ composite. A human operator produced a good composite of an unfamiliar target face, which was in view throughout the process, after viewing 162 faces, over 27 iterations and took approximately 20 minutes (see Figure 1). A recognisable composite of Tony Blair was produced from memory after 23 iterations, viewing 138 faces (see Figure 2). Both of these composites were constructed using the SMM genetic algorithm. Formal human experimental evaluation of the Eigen-fit system is currently in progress.



*Figure 1: SMM human trial with target face visible: a) Starting face, b) and c) are intermediate points in the evolutionary process, d) final generated composite after 27 iterations (162 faces viewed), e) the actual target face.*



*Figure 2: SMM human trial for famous face from memory: a) Starting face, b) and c) are intermediate points in the evolutionary process, d) final generated composite after 23 iterations (138 faces viewed), e) addition of hair to facial composite.*

Frowd, Hancock and Carson (2004) found that naïve judges could name 10% of Evo-fit composites of celebrities produced from memory, compared to 17% of composites produced by an E-fit operator. The poorer performance of Evo-fit could have been attributable to the age range of the celebrities being inappropriate to the database used to generate the PCA space for Evo-fit. The age range of celebrities was appropriately restricted in a second experiment, in which the target faces were visible during the production of the composite. The naming rate of Evo-fit composites was 25% under these conditions, which is similar to comparable data for E-fit.

Frowd et al (in press) evaluated the utility of Evo-fit, E-fit, Profit, FACES and a police sketch artist under more forensically realistic conditions. The ‘witness’ viewed a target face of a celebrity. The celebrities were not very famous and were chosen to be unfamiliar to each witness. After a two-day delay, each witness underwent a cognitive interview and worked with an appropriately trained operator to construct a composite. The utility of the composites was evaluated by three groups of participants each given one of three tasks: naming, sorting, and identification. The sorting task required participants to match composites to the appropriate face from an array of all the targets. The identification task required participants to match the composite to the target face from a line-up including distracters chosen to be similar in appearance. The naming rate was very low, even when conditionalised by the number of participants who were familiar with the target celebrities. The naming rate of sketches (8.1%) was significantly higher than for PROfit (1.3%) and E-fit (0%), but did not achieve a statistically significant difference compared to Evo-fit (3.6%) and FACES (3.2%). The sorting task produced a much higher level of performance but a similar result. Performance was significantly better for the sketches (54%) than for Evo-fit (39%) and the other systems (25% - 42%). There was no significant difference between the performance of any of the composite systems. E-fit performed best in the identification task (60% compared to 47% for sketches and 31% for Evo-fit). Performance with E-fit was significantly better than all other systems except sketches. However, performance on the identification task was not correlated with performance on naming. In contrast, sorting performance showed a significant correlation with naming rate. Naming is usually considered to be the most forensically relevant test, therefore the lack of an association between ‘identification’ performance and naming suggests that the identification task should be interpreted with caution.

The use of genetic algorithms is an exciting development, which exploits contemporary theory of face processing. The GA technique can perform at similar levels to the current composite systems, it remains to be demonstrated whether they



will prove more effective than current composite methods. Like all systems, the GA methods incorporate certain psychological assumptions about memory for faces that deserve to be more rigorously evaluated. First, research suggests that for Caucasian faces, hair is the single most salient cue for witnesses (Ellis, 1986). While hairstyle is selected at an early stage in some systems, it is divorced from the choice of Eigenfaces. Second, many systems require witnesses to grade the similarity of faces, but earlier research suggests that perhaps only half the composites produced by witnesses are of an appropriate physiognomic type (Christie et al., 1981) and that witnesses are also poor at making absolute judgements of similarity with any degree of accuracy from memory (Clark, 2000). There is also the danger that viewing approximate likenesses may interfere with memory for the original face (Turner, 2000). Moreover, a skilled police artist can still outperform all current systems that have been evaluated so far.

## CONCLUSIONS

The skilled police artists remain the benchmark against which all systems must compare and no mechanical or software system has yet to equal or outperform them. However, while artists are quick to trumpet their successes, they have also had their failures and the overall level of accuracy is hard to compute for a skill so idiosyncratic and poorly understood. After three decades of intensive research, it is still unclear for any technique what predicts or postdicts a successful interview. Witnesses are inconsistent in the quality of composites they reproduce from one face to another and over time (Davies et al., 1978a). Neither the witnesses themselves nor the operators are effective in estimating when a likeness is likely to prove to be of good or poor quality (Kovera et al., 1997). A good likeness appears to depend upon an elusive combination of a face whose features may be readily reproduced; an observant and articulate witness, and a skilled operator who knows how to ask the right questions (Davies et al., 1983).

This is not to deny the progress that has been achieved through research and development. Some of the more obvious sources of error evident in earlier systems have been identified and removed. These include a lack of relevant features and sufficient flexibility of size and positioning to model the full range of faces. For the male Caucasian face, most software systems now allow the skilled operator to fashion a recognisable likeness from life or a photograph (Brace et al., 2000; Cutler et al., 1988). Likewise, fourth generation systems permit witnesses to work on total faces rather than the traditional approach emphasising individual features (Gibson et al., 2003).

One area of continuing controversy concerns the possible inhibiting effect of verbal description on facial recall. Dodson, Johnson and Schooler (1997) demonstrated experimentally that recognition for faces can be impaired if the observer is required to verbally describe them prior to recognition: the 'verbal overshadowing effect.' It has been speculated that to provide detailed verbal descriptions might impair witness's ability to subsequently select appropriate features. Clark (2000) for instance, reported that for E-fit, the recommended practice of re-interviewing the witness about the suspect's appearance midway through construction had a detrimental effect upon final composite quality, a finding consistent with overshadowing. However, verbal overshadowing is not an inevitable consequence of describing a face, even under laboratory conditions (Meissner & Brigham, 2001) and delay serves to reduce any potential impairment (Finger & Pezdek, 1999). The

conditions under which verbal encoding interferes with facial memory remain poorly understood. The retrieval based interference explanation assumes that verbalisation impairs the original memory trace of the face (Meissner, Brigham & Kelley, 2001). However, in some circumstances it appears that verbal recall and visual recognition processes function independently (Davies, 1986a) and an explanation of the verbal overshadowing effect in terms of a criterion shift seems at least as plausible (Clare & Lewandowski, 2004).

One consideration which perhaps has been insufficiently challenged is the belief that memory for a briefly observed and unfamiliar face is sufficiently detailed to construct a successful composite. This belief appears to be based on the frequently iterated statement that face recognition is far superior to face recall and our ability to recognise faces, often after many years, testifies to a robust and unique encoding system for all faces. More recent research on face recognition however suggests that familiar and unfamiliar faces are processed in different ways which result in striking differences in ease of recognition (Bruce & Young, 1998). Even degraded images of familiar individuals caught on CCTV are readily recognised (Burton, Wilson, Cowan & Bruce, 1999), but, unfamiliar faces seen on CCTV are matched to an appropriate photograph very inaccurately indeed, even when participants have continuous access to an image of the face as they carry out the task (Bruce, Henderson, Newman, & Burton, 2001; Davies & Thasen, 2000; Kemp, Towell & Pike, 1997).

Research from other areas of face processing suggests that memory for the appearance of novel faces may be fragmentary and inadequate. Ellis (1984) noted that verbal descriptions, both in the presence of the face and from memory, were selective and incomplete. Even in recognition memory for novel faces, faces which share certain dominant attributes such as hair style and face shape are readily confused (Davies, Shepherd & Ellis, 1979). Learning a face takes time and repeated exposure under different viewing conditions (Bruce, 2003).

Schema theory has demonstrated that where memory is imperfect, then plausible reconstruction is likely to take place, which may or may not be accurate (Brewer, 1996). In a task like constructing a face, which requires exhaustive recall of all features, there are opportunities for attitudes and assumptions to fill gaps and colour the constructive process. Some years ago, Shepherd, Ellis, McMurrin and Davies (1978) demonstrated the impact of negative and positive stereotypes had on Photofit reconstructions. Witnesses constructed composites which were judged as more intelligent and handsome when they were told the man was a lifeboat captain than when he was described as a murderer (see also Oliver, Jackson, Moses & Dangerfield, 2004 for an example of the influence of racial stereotyping on face recall). More recently, Davies and Oldman (1999) replicated Shepherd et al's finding using familiar faces and showed that attitudes also influenced quality of likeness. Faces made by persons who disliked the target were of a better quality than those made by persons who liked them. As the authors observed: contempt appears to breed familiarity.

It seems likely that the largest distortions due to affect and stereotyping will occur on unfamiliar faces viewed for fleeting periods; often the conditions when witnesses to crime view actual suspects. In these circumstances, it may be that for many witnesses, composite production imposes an unrealistic burden upon them, with inevitable consequences for composite quality, irrespective of the system employed. Perhaps, in the light of recent findings, composite production should be reserved for witnesses who have had extensive experience of the person concerned. Perhaps feature selection should be confined to items mentioned by witnesses in their verbal

descriptions. Intelligent systems could be developed that could accurately ‘suggest’ missing features from existing choices of other parts of the face, rather than relying upon guesses fuelled by feelings and stereotypes.

Probably the first encounter between psychologists and the Identikit was described by Connolly and McKeller (1963): “Having seen this device, and having been subjects in a demonstration, we consider this to be a marked improvement [over verbal descriptions] but also a ‘psychological Pandora’s box’” (p.22) adding that “the problem of identification would repay psychological enquiry” (p.23). Four generations of composite systems have now been reviewed together with the psychological enquiry they have provoked. While measurable progress has been made and all systems may claim successes, perhaps the quest for the perfect system may be illusory and we must learn to live within the limitations of witness memory.

#### REFERENCES

- Allison, H. C. (1973). Personal identification. Boston, MA: Holbrook Press
- Aspley Limited (1993). E-fit. Hatfield, United Kingdom: Aspley Limited.
- Brewer, M. B. (1996). When stereotypes lead to stereotyping: The use of stereotypes in person perception. In C. N. Macrae, C. Stangor & M. Hewstone (Eds.), Stereotypes and stereotyping (pp. 254-275). New York: Guilford.
- Bruce, V. (2003, June). Getting to know you – How we learn new faces. Final report to the Economic and Social Research Council. London:ESRC
- Bruce, V., & Valentine, T. (1988). When a nod’s as good as a wink: The role of dynamic information in facial recognition. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory: Current research and issues. Vol. 1 (pp.169-174). Chichester, UK: Wiley.
- Bruce, V. & Young, A. (1998). In the eye of the beholder: The science of face perception. Oxford: Oxford University Press
- Bennett, P. (1986). Face recall: A police perspective. Human Learning, 5, 197-202.
- Benson, P. J. & Perrett, D. L. (1991). Perception and recognition of photographic quality caricatures: Implications for the recognition of natural images. European Journal of Cognitive Psychology, 3, 103-135.
- Boylan, J. (2000) Portraits of guilt: The woman who profiles the faces of America’s deadliest criminals. New York: Pocket books
- Brace, N. A., Pyke, G. E., & Kemp, R. I. (2000). Investigating E-fit using famous faces. In A. Czerederecka, T. Jaskiewicz-Obydzinska & J. Wojcikiewicz (Eds.), Forensic psychology and law : Traditional questions and new ideas (pp. 272-276). Krakow, Poland: Institute of Forensic Research Publishers.
- Brace, N. A., Pyke, G. E., Kemp, R. I., Turner, J. & Bennett, P. (2001) Does the presentation of multiple facial composites improve suspect identification? Unpublished paper, Department of Psychology, The Open University.
- Bruce, V. (2003). Getting to know you-How we learn new faces. Final report to the Economic and Social Research Council. Swindon: ESRC
- Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. ((192). The importance of ‘mass’ in the line drawings of faces. Applied Cognitive Psychology, 6, 619-628.

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

Bruce, V., Henderson, Z., Newman, C. & Burton, A.M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. Journal of Experimental Psychology: Applied, 7, 207-218.

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

Byatt, G. & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: implications for models of face recognition. Vision Research, 38, 2455-2468.

Chiroro, P. & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. Quarterly Journal of Experimental Psychology, 48A, 879-894.

Clare, J. & Lewandowsky, S. (2004). Verbalising facial memory: Criterion effects in verbal overshadowing. Journal of Experimental Psychology: Learning, Memory and Cognition, 30, 739-755.

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

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

Christie, D., & Ellis H. (1981). Photofit constructions versus verbal descriptions of faces. Journal of Applied Psychology, 66, 358-363.

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

Connolly, K. & McKeller, P. (1963). Forensic psychology. Bulletin of the British Psychological Society, 16, 16-24.

Cootes, T.F., Edwards, G. J., Taylor, C. J. (1998). Active appearance models. In H. Burkhardt & B. Neumann (eds.) Proceeding of the European Conference on Computer Vision. Volume 2. (pp. 484-498.) Springer

Cootes, T.F., & Taylor, C. J. (2001). Statistical models of appearance for medical image analysis and computer vision. Proceedings of SPIE Medical Imaging, 236-248.

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

Craw, I. & Cameron, P. (1991). Parametising images for recognition and reconstruction. In: P. Mowforth (ed.) Proceedings of the British Machine Vision Conference 1991. (pp. 367-370). New York: Turing Institute Press and Springer-Verlag.

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

Davies G. (1981). Face recall systems. In G. Davies, H. Ellis, & J. Shepherd (Eds.), Perceiving and remembering faces (pp. 227-250). London: Academic Press.

Davies, G. M. (1986a). The recall and reconstruction of faces: Implications for theory and practice. In H. D. Ellis, M. A. Jeeves & A. Young (Eds.), Aspects of face processing (pp. 388-398). Dordrecht, The Netherlands: Nijhoff.

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

- Davies, G. M. (1996). Children's identification evidence. In S.L. Sporer, R.S. Malpass and G. Koehnken (Eds.), Psychological issues in eyewitness identification (pp.233-258). Mahwah, NJ: Erlbaum.
- Davies, G., & Christie, D. (1982). Face recall: an examination of some factors limiting composite production accuracy. Journal of Applied Psychology, *67*, 103-109.
- Davies G., Ellis, H., & Shepherd, J. (1978a). Face identification. The influence of delay upon accuracy of Photofit construction. Journal of Police Science and Administration, *6*, 35-42.
- Davies G., Ellis, H., & Shepherd, J. (1978b). Face recognition accuracy as a function of mode of representation. Journal of Applied Psychology, *63*, 180- 187.
- Davies, G., Ellis, H., & Shepherd J. (1981). Perceiving and remembering faces. London: Academic Press.
- Davies, G. M., Ellis, H. D., & Shepherd, J. W. (1985, May 16). Wanted-Faces that fit the bill. New Scientist, No 1456, pp. 26-29.
- Davies, G. & Little, M (1990). Drawing on memory: Exploring the expertise of the police artist. Medicine. Science and the Law, *30*, 345-353.
- Davies, G., & Milne, A. (1985). Eyewitness composite production. A function of mental or physical reinstatement of context. Criminal Justice and Behavior, *12*, 209-222.
- Davies, G., Milne, A., & Shepherd, J. (1983). Searching for operator skills in face composite reproduction. Journal of Police Science and Administration, *11*, 405-409.
- Davies, G., & Oldman, H. (1999). The impact of character attribution on composite production: A real world effect? Current Psychology, *18*, 128-139.
- Davies, G. M., Shepherd, J. W., & Ellis, H. D. (1979). Similarity effects in face recognition. American Journal of Psychology, *92*, 507-523.
- Davies, G., Shepherd, J. W., Shepherd, J., Flin, R., & Ellis, H. (1986). Training skills in police Photofit operators. Policing, *2*, 35-46.
- Davies, G. & Thasen, S. (2000). Closed-circuit television: How effective an identification aid? British Journal of Psychology, *91*, 411-426
- Davies, G.M., van der Willik, P. & Morrison, L. (2000). Facial composite production: A comparison of mechanical and computer-driven systems. Journal of Applied Psychology, *85*, 119-124.
- De Haan, M., Humphreys, K. & Johnson, M. (2002). Developing a brain specialized for face perception: A converging methods approach. Developmental Psychobiology, *40*, 200-212.
- Domingo, F. (1984, June). Composite art: The need for standardization. Identification News, pp.7-15.
- Ellis, H. D. (1984). Practical aspects of face memory. In G. Wells and E. Loftus (Eds.), Eyewitness testimony (pp. 12-37). Cambridge: Cambridge University Press.
- Ellis, H. (1986). Face recall: a psychological perspective. Human Learning, *5*, 189-196.
- Ellis, H., Davies, G., & McMurrin, M. (1979). Recall of white and black faces by white and black witnesses using the Photofit system. Human Factors, *21*, 55-59.
- Ellis, H., Davies, G., & Shepherd, J. (1976). An investigation of the Photofit system for recalling faces (Final report, Grant No. HR 3123/1). Swindon: Social Science Research Council.
- Ellis, H., Davies, G., & Shepherd, J. (1978x). A critical examination of the Photofit system for recalling faces. Ergonomics, *21*, 297-307.

- Ellis, H., Davies, G., & Shepherd J. (1978b). Remembering pictures of real and unreal faces: some practical and theoretical considerations. British Journal of Psychology, *69*, 467-1174.
- Ellis, H., Shepherd, J., & Davies, G. (1975). An investigation of the use of the Photofit technique for recalling faces. British Journal of Psychology, *66*, 29-37
- Finger, K. & Pezdek, K. (1999). The effect of the cognitive interview on face identification accuracy: Release from verbal overshadowing. Journal of Applied Psychology, *84*, 340348.
- Flin, R., Markham, R. & Davies, G. M.( 1988). Making faces: Developmental trends in the construction and recognition of face composites. Journal of Applied Developmental Psychology, *10*, 123-137.
- Frowd, C., Hancock, P. J. B. & Carson, D. (2004). EvoFIT: A holistic evolutionary facial imaging technique for creating composites. Association for Computing Machinery Transactions on Applied Psychology , *1*, 1-21.
- Frowd, C., Carson, D., Ness, H., McQuiston-Surrett, D., Richardson, J. Baldwin, H & Hancock, P. (in press). Contemporary composite techniques: The impact of a forensically-relevant target delay. Legal and Criminological Psychology
- Garcia, E. & Pyke, C. (1977). Portraits of crime, New York: Condor.
- Gibling, F. & Bennett, P. (1994). Artistic enhancement in the production of Photofit likenesses; An examination of its effectiveness in leading to suspect identification. Psychology, Crime and the Law, *1*, 93-100.
- Gibson, S. Pallares Bejarano, A. & Solomon, C. (2003). Synthesis of photographic quality facial composites using evolutionary algorithms. In: R. Harvey & J. A. Bangham (eds.) Proceedings of the British Machine Vision Conference 2003. (pp. 221-230). London: British Machine Vision Association
- Gillenson, M., & Chandrasekaren, B. (1975). A heuristic strategy for developing human facial images on a CRT. Pattern Recognition, *7*, 187-196.
- Green, D. L., & Geiselman, R. E. (1989). Building composite facial images: Effect of feature saliency and delay of construction. Journal of Applied Psychology, *74*, 714-721.
- Haig, N. D. (1986). Investigating face recognition with an image processing computer. In H. D. Ellis, M. A. Jeeves & A. Young (Eds.), Aspects of face processing (pp. 410-425). Dordrecht, The Netherlands: Nijhoff.
- Hancock, P. J. B. (2000). Evolving faces from principal components. Behaviour Research methods, Instruments and Computers, *32*, 327-333.
- Homa, G. (1983). The law enforcement composite sketch artist. NJ: Author.
- Jackson, R. L. (1967). Occupied with crime. London: Harrap.
- Kemp, R., Towell, N. & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. Applied Cognitive Psychology, *11*, 211-222.
- King, D. (1971). The use of Photofit 1970-1971: A progress report. Police Research Bulletin, no. 18, pp.40-44.
- Kitson, A., Darnbrough, M., & Shields, E. (1978). Let's face it. Police Research Bulletin, no. 30, pp. 7-13.
- Koehn, C., & Fisher, R. P. (1997). Constructing facial composites with the Mac-a-Mug Pro system. Psychology, Crime and Law, *3*, 209-218.
- Kovera, M. B., Penrod, S., Pappas, C., & Thill, D. (1997). Identification of computer-generated facial composites. Journal of Applied Psychology, *82*, 235-246.

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

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

Laughery, K. & Smith, V. L. (1978). Suspect identification following exposure to sketches and Identikit composites. Proceedings of the Human Factors Society, 22<sup>nd</sup> Annual Meeting, Detroit (pp.631-635).

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

Levi, A. M. (1997). Police composites: Do they contribute to convictions? Unpublished manuscript: Division of Identification and Police Science, Israeli Police Headquarters, Jerusalem.

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

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

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

Oliver, M. B., Jackson, R. L., Moses, N. N., & Dangerfield, C. L. (2004). The face of crime: Viewers' memory of race-related facial features of individuals pictured in the news. Journal of Communication, 54, 88-104.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A. & Valentin, D. (1995). A perceptual learning theory of the information in faces. In T. Valentine (ed.) Cognitive and computational aspects of face recognition: Explorations in face space (pp.159-182). London: Routledge.

Owens, C. (1970, November). Identikit enters its second decade-ever growing at home and abroad. Finger Print and Identification Magazine, pp. 3-8 & 11-17.

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

Penserga, B.(2003, October 20). Police sketch artists yield to computer composites. The Daily Times (Delaware),

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

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

Rhodes, G. (1996). Superportraits: Caricatures and recognition. Psychology Press. Hove.

Schwartz-Kenney, Norton, C., Chalkley, B., Jewett, J., & Davis, K. (1996, February). Building a composite of a stranger: Young children's use of the Identi-Kit. Paper presented at the Biennial Conference of the American Psychology-Law Society, Hilton Head, NC.

Sergent, J. (1984). An investigation into component and configurational processes underlying face perception. British Journal of Psychology, 75, 221-242.

Shaherazam (1986). The Mac-a-Mug pro Manual. Milwaukee, WI: Shaherazam.

- Shepherd, J., Davies, G., & Ellis, H. (1981). Studies of cue saliency. In G. Davies, H. Ellis, & J. Shepherd (Eds.), Perceiving and remembering faces (pp.105-132). London: Academic Press.
- Shepherd, J. W., & Ellis, H. D. (1996). Face recall-Methods and problems. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), Psychological issues in eyewitness identification (pp. 87-116). Mahwah, NJ: Erlbaum.
- Shepherd, J. W., Ellis, H. D., McMurrin, M., & Davies, G. M. (1978). Effect of character attribution on Photofit construction of a face. European Journal of Social Psychology 8, 263-268.
- Sirovich, L & Kirby, M. (1987). Low dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A, 4, 519-524.
- Sondern, F. (1964, April). The box that catches criminals. Readers' Digest, pp.37-44.
- Taylor, K.T. (2001). Forensic art and illustration. Boca Raton, FL: CRC Press.
- Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3, 71-86.
- Tredoux, C. (2001). A non-verbal approach to face reconstruction by eyewitnesses. Report on National Research Foundation CSD grant 15/1/3/16/0158
- Tredoux, C. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. Journal of Experimental Psychology: Applied, 8, 180-193.
- Tredoux, C., Rosenthal, Y., NunezD. & da Costa, L. (1999). Face reconstruction using a configural, eigenface-based composite system. Paper presented to the third meeting of the Society for Applied Research Memory and Cognition, Boulder, Colorado, July 1999. <http://web.uct.ac.za/depts/psychology/plato/> Retrieved 5/12/03.
- Tanaka, J.W. & Farah, M. J. (2003). The holistic representation of faces. In: Peterson, M.A. & Rhodes, G. Perception of faces, objects and scenes. Oxford University Press, Oxford. Pp. 53-74.
- Valentine, T. (1991a). A unified account of the effects of distinctiveness, inversion and race in face recognition. Quarterly Journal of Experimental Psychology, 43A, 161-204.
- Valentine, T. (1991b). Representation and process in face recognition. In: R. Watt (ed.) Pattern recognition by Man and Machine. (Vol. 14 in 'Vision and Visual Dysfunction' series edited by J. Cronly-Dillon) London: Macmillan Press. **Page nos.**
- Valentine, T., (1995). Cognitive and computational aspects of face recognition: Explorations in face space. London: Routledge.
- Valentine, T. (2001). Face-space models of face recognition. In: M.J. Wenger & J.T. Townsend (eds.) Computational, geometric, and process perspectives on facial cognition: Contexts and challenges (pp. 83-113). Mahwah: LEA..
- Valentine, T. & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. Quarterly Journal of Experimental Psychology, 44A, 671-703.
- Venner, B. R. H. (1969). Facial identification techniques. Police Research Bulletin, no.13, pp17-20.
- Wells, G., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. Memory and Cognition, 12, 338-344.
- Wogalter, M., & Marwitz, D. (1991). Face composite construction: In-view and from-memory quality and improvement with practice. Ergonomics, 34, 459-468.