

## Central Lancashire Online Knowledge (CLoK)

Title	Recovering faces from memory: the distracting influence of external facial features
Type	Article
URL	<a href="https://clock.uclan.ac.uk/id/eprint/2211/">https://clock.uclan.ac.uk/id/eprint/2211/</a>
DOI	<a href="https://doi.org/10.1037/a0027393">https://doi.org/10.1037/a0027393</a>
Date	2012
Citation	Frowd, Charlie, Skelton, Faye Collette, Atherton, Chris J, Pitchford, Melanie, Hepton, Gemma, Holden, Laura, McIntyre, Alex H. and Hancock, Peter J. B. (2012) Recovering faces from memory: the distracting influence of external facial features. <i>Journal of Experimental Psychology: Applied</i> , 18 (2). pp. 224-238. ISSN 1939-2192
Creators	Frowd, Charlie, Skelton, Faye Collette, Atherton, Chris J, Pitchford, Melanie, Hepton, Gemma, Holden, Laura, McIntyre, Alex H. and Hancock, Peter J. B.

It is advisable to refer to the publisher's version if you intend to cite from the work.  
<https://doi.org/10.1037/a0027393>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

# Recovering faces from memory: the distracting influence of external facial features

Charlie D. Frowd (1\*)

Faye Skelton (1)

Chris Atherton (1)

Melanie Pitchford (2)

Gemma Hepton (1)

Laura Holden (1)

Alex H. McIntyre (3)

Peter J.B. Hancock (3)

(1) School of Psychology, University of Central Lancashire, PR1 2HE

(2) Department of Psychology, Lancaster University, LA1 4YF

(3) Psychology, School of Natural Sciences, University of Stirling, FK9 4LA

\*Corresponding author: Charlie Frowd, School of Psychology, University of Central Lancashire, Preston PR1 2HE, UK. cfrowd@uclan.ac.uk. (01772) 893439.

Running head: External facial context for face construction

## Abstract

Recognition memory for unfamiliar faces is facilitated when contextual cues (e.g. head pose, background environment, hair and clothing) are consistent between study and test. By contrast, inconsistencies in external features, especially hair, promote errors in unfamiliar face-matching tasks. For the construction of facial composites, as carried out by witnesses and victims of crime, the role of external features (hair, ears and neck) is less clear, although research does suggest their involvement. Here, over three experiments, we investigate the impact of external features for recovering facial memories using a modern, recognition-based composite system, EvoFIT. Participant-constructors inspected an unfamiliar target face and, one day later, repeatedly selected items from arrays of whole faces, with 'breeding', to 'evolve' a composite with EvoFIT; further participants (evaluators) named the resulting composites. In Experiment 1, the important internal-features (eyes, brows, nose and mouth) were constructed more identifiably when the visual presence of external features was decreased by Gaussian blur during construction: higher blur yielded more identifiable internal-features. In Experiment 2, increasing the visible extent of external features (to match the target's) in the presented face-arrays also improved internal-features quality, although less so than when external features were masked throughout construction. Experiment 3 demonstrated that masking external-features promoted substantially more identifiable images than using the previous method of blurring external-features. Overall, the research indicates that external features are a distractive rather than a beneficial cue for face construction; the results also provide a much better method to construct composites, one that should dramatically increase identification of offenders.

(250 words)

There is a wealth of evidence to suggest that face recognition is a holistic process. For example, recognition of an individual facial feature (e.g., eye, nose or mouth) is more accurate when seen in the context of a complete face, compared with when it is seen as an isolated part (e.g. Davies & Christie, 1982; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). In addition, instructions to participants that promote encoding of whole faces, such as personality attribution, have been shown to facilitate recognition, compared with instructions that focus on individual parts—such as length of nose or spacing between the eyes (e.g. Berman & Cutler, 1998; Shapiro & Penrod, 1986; Sporer, 1991). Similarly, manipulations that promote global over local face processing at recognition also improve performance (e.g. Berman & Cutler, 1998; Macrae & Lewis, 2002), while those that do the opposite decrease it (e.g. Meissner, Sporer, & Susa, 2008). Inverting a face interferes with processing of distances between individual features, the relational information, and this manipulation illustrates the importance of seeing faces in an upright context (e.g. Leder & Carbon, 2006; Thompson, 1980). Other research reveals that holistic processing is so strong as to make identification difficult when top and bottom halves of different faces are seen as a single intact item, unless the halves are presented misaligned (Hole, 1994; Young, Hellawell, & Hay, 1987). Also, transformations such as facial motion and facial caricature affect an entire image and similarly indicate the holistic nature of face recognition (e.g. Lander, Christie, & Bruce, 1999; Lee & Perrett, 2000).

Holistic processing is qualitatively different for faces seen on many occasions, familiar faces, and for faces we have seen infrequently, or even just once, so-called unfamiliar faces (e.g. Hancock, Bruce, & Burton, 2000). For the latter, unfamiliar case, recognition accuracy is modulated by the context in which the face is perceived (for a recent review, see Johnston & Edmonds, 2009). For this reason, any change in facial appearance, such as angle-of-view and facial expression, hampers recognition when familiarity is low (e.g. Davies & Milne, 1982; Hill & Bruce, 1996). These effects can also be additive, with multiple changes further degrading recognition (e.g. Bruce, 1982). Other research has revealed that recognition is a function of the place in which a face is seen—that is, the background setting (environmental context). Memon and Bruce (1983) found that unfamiliar face recognition was worst when the background was completely different (e.g. face seen in restaurant, then in bank), better when contexts were semantically-similar (e.g. different restaurants) and best with identical context (e.g. same restaurant). Sporer (1993) discovered that clothing likewise acts as a retrieval cue. In his study, recognition memory was improved when the context provided by clothes were visible on a person's shoulders compared with when they had been concealed. Context effects like these are often interpreted in terms of the Tulving and Thomson (1973) encoding specificity theory: recognition is facilitated if the cues present at test overlap with those seen at encoding.

For unfamiliar faces, cognitive processing is dependent on both internal features

(eyes, brows, nose and mouth) and external features (hair, forehead, ears and neck). Specifically, external features provide a context that improves our ability to recognise (e.g. Ellis, Shepherd, & Davies, 1979; Endo, Takahashi, & Maruyama, 1984) and to match (Bruce et al., 1999) internal features. The role of hair, an external feature, would appear to be especially important: hair is responsible for a high proportion of false identifications in unfamiliar face-matching tasks (Bruce et al., 1999; Frowd, Bruce, McIntyre, & Hancock, 2007) and is one of the three ‘cardinal’ features—along with age and face shape—that we use to classify faces (Ellis, 1986). Hair also functions as a useful retrieval cue, similar to cues provided by clothes and background, since concealing hair at test decreases unfamiliar face recognition accuracy (Cutler, Penrod, O'Rourke, & Martens, 1987).

In contrast to our processing of unfamiliar faces, repeated exposure to a specific face—that is, the acquisition of familiarity—produces a type of processing that is largely invariant to cues like head pose, expression, background and external features. Such recognition of familiar faces relies more on internal than external features (e.g. Andrews, Davies-Thompson, Kingstone, & Young, 2010; Campbell et al., 1999; de Haan & Hay, 1986; Ellis et al., 1979; Young et al., 1985), is highly accurate (e.g., Bruce, 1986), and is associated with higher levels of confidence than recognition of unfamiliar faces (e.g. Burton, Wilson, Cowan, & Bruce, 1999). Familiar face recognition is also robust, even following long intervals between successive exposures (Bruck, Cavanagh, & Ceci, 1991) or when stimuli have been degraded—for example following coarse image pixelation (e.g. Lander, Bruce, & Hill, 2001) or moderate levels of Gaussian blur (e.g. Hole, George, Eaves, & Rasek, 2002).

Both types of face familiarity are engaged in the construction and recognition facial composites. In a forensic setting, facial composites are images of alleged criminal offenders, produced from memory by witnesses (who may also be victims). The traditional method to produce them is for the witness to describe the offender's face and then to construct it by selecting from a large range of individual facial features (eyes, nose, hair, mouth, etc). The resulting image is then shown to other people, often members of the public, with the aim that someone will recognise the face and assist an investigation by providing police with a name. Composite construction therefore involves unfamiliar face processing, as the face is normally unknown to the witness, whereas composite recognition involves familiar face processing, naming.

Modern ‘feature’ methods of face construction (e.g. E-FIT and PRO-fit) involve witnesses selecting individual features in the context of a complete face, rather than the old method of looking at isolated elements, as was the case with Photofit and Identikit systems—an approach that should, according to research, promote a more identifiable image (e.g. Davies & Christie, 1982; Homa, Haver, & Schwartz, 1976; Tanaka & Farah, 1993), and for which we have recently found further support

(Skelton, Frowd, & Speers, 2011). Unfortunately, even using this enhanced method of feature selection, constructors rarely produce highly-recognisable faces (e.g. Brace, Pike, & Kemp, 2000; Bruce, Ness, Hancock, Newman, & Rarity, 2002; Davies, van der Willik, & Morrison, 2000; Frowd, Carson, Ness, Richardson, et al., 2005), especially when the delay between encoding and construction is the one day or two days typical of police investigations (e.g. Frowd, Carson, Ness, McIntyre, et al., 2005; Frowd, Bruce, Ness, et al., 2007; Frowd, Pitchford, et al., 2010). Describing and selecting individual facial features are not tasks at which humans excel (e.g. Davies, Shepherd, & Ellis, 1978; Ellis, Shepherd, & Davies, 1980; Frowd, Carson, Ness, McIntyre, et al., 2005; Woodhead & Baddeley, 1984), and so constructed likenesses tend to be of poor quality (*ibid.*).

It is for this reason that researchers have sought new methods to recover faces from memory (for examples, see Frowd, Bruce, & Hancock, 2008, or Frowd, *in press*). The most promising method does not necessitate that witnesses produce a facial description nor select individual features. Instead, computer software synthesizes complete faces with random characteristics (for a description of how to do this, typically achieved via Principal Components Analysis, or PCA, see Craw & Cameron, 1991; Frowd, Hancock, & Carson, 2004; or Hancock, 2000). These unfamiliar identities are presented in simultaneous face arrays, and witnesses select items that bear an overall resemblance to an offender. Evolutionary Algorithms (EAs) are then employed to combine characteristics of the selected items, producing alternatives for further selection. After a few iterations, the face arrays converge on a specific identity, which is ideally an identifiable likeness of the target (offender). In essence, the person constructing the face provides feedback (by selecting items from arrays) to guide a search through face space (see Mitchell, 1996, for an overview of EAs, and Lewis, 2004, for a review of face space).

We are aware of three similar methods that follow these evolutionary principles (Frowd et al., 2004; Gibson, Solomon, Maylin, & Clark, 2009; Tredoux, Nunez, Oxtoby, & Prag, 2006). In our EvoFIT method, witnesses choose the set of external features that best matches their memory of the target, then select from arrays which vary first by feature shape and relational information, and second by colouring of individual features and overall skin tone, following which new arrays are ‘bred’ for further selection. Software tools allow final adjustment and placement of feature shape, and the addition of shading, wrinkles and other textural artwork.

Two main developments for EvoFIT have facilitated recovery of a face from a composite constructor’s memory. First, the visibility of external features in the face arrays is reduced by applying Gaussian blur to them during face selection (Frowd, Pitchford, et al., 2008). Since unfamiliar faces are presented, the idea was that this image filtering would facilitate selection of the internal features—the region that is mainly responsible for successful naming of the composite by another person (e.g.

Ellis et al., 1979; Frowd, Skelton, Butt, Hassan, & Fields, 2011). Arguably due to the fallibility of memory, even with this enhancement, constructors still tend to evolve faces that include appreciable error. So, as a second development, methods were designed to allow users to rectify inaccuracies in an evolved face (Frowd, Bruce, McIntyre, et al., 2006). These ‘holistic tools’ include attractiveness, honesty, masculinity, health and extroversion. In use, constructors see their evolved face manipulated along each holistic scale and locate the setting which they believe to yield the best likeness.

Frowd, Pitchford, et al. (2010) assessed the effectiveness of these two developments. Participants inspected an unfamiliar target face—then, two days later, described and constructed a composite of it. External-features blurring and holistic tool use were somewhat-equally effective on their own; when used together, the faces produced were named by other people with a mean of 24% correct. Naming was only 4% correct for faces produced by the traditional method of selection by individual features.

In the current work, we investigate further the influence of exterior facial context on constructing faces. In Frowd, Pitchford, et al. (2010), participants constructed much more identifiable images from face arrays seen with external-features blur than when these features were fully visible (Cohen’s  $d = 1.2$ ). The level of blurring was such as to render recognition difficult if extended across the entire image (e.g. Thomas & Jordan, 2002)—for an example, see Figure 1, far right. Blurring may help to reduce some or all of the negative influence of external features on accurate face construction (e.g. Bruce et al., 1999). Alternatively, it is conceivable that blurring may also interfere with a potentially useful retrieval cue, the external features themselves (e.g. Ellis et al., 1979). If this is the case, then less blur may be more valuable—having the same effect as maintaining accurate or similar backgrounds between study and test (e.g. Davies & Milne, 1982; Memon & Bruce, 1983), and when hair, clothing or both are visible at recognition (Cutler et al., 1986; Sporer, 1993). So, if the latter explanation is true, then reducing the level of blurring will allow greater positive influence of external features and promote a more identifiable likeness than for composites made under ‘high’ blur; if the former is true, then the opposite result will be observed: less intense blurring will allow external features to exert greater negative effect, yielding a worse constructed likeness. These competing hypotheses were tested in the first experiment. Experiment 2 used a similar rationale to explore the role of external features during face construction, not by blurring, but by increasing the extent to which they matched the target. These two experiments together suggest that external features exert a negative influence on the face construction process. Experiment 3 then explored the potential advantage of masking external features while internal features are being constructed relative to the current EvoFIT procedure whereby external features are blurred during face selection and

breeding.

## **Experiment 1**

Gaussian blur is an image filtering technique often measured in cycles per face-width (f/w) (e.g. Costen, Parker, & Craw, 1996; Thomas & Jordan, 2002), but it can also be measured as a radius (pixels) (e.g. Jeong, Kim, & Lee, 2010). In both cases, higher values degrade an image to a greater extent. In Frowd, Pitchford, et al. (2010), external features were blurred at 8 cycles per f/w. Here, this ‘high’ blur was compared against two evenly spaced intermediate levels, ‘medium’ (6 cycles per f/w) and ‘low’ (4 cycles per f/w); a fourth condition was included with blurring disabled throughout face construction. In Stage 1, below, participant-constructors were randomly assigned to one of four levels of external-features blur (none, low, medium or high). Each person evolved a single composite image from memory using EvoFIT with face arrays set to present external features at this level of blur. Figure 1 illustrates the four levels used.

Figure 1 about here

Composites were evaluated in Stage 2, below, by asking other people (composite-evaluators) to name them, as per police procedure. We predicted that faces evolved with high blur would be named more accurately than composites with no blur at all, as found previously (Frowd, Park, et al., 2008; Frowd, Pitchford, et al., 2010). However, construction using intermediate (low, medium) blur, relative to no-blur and high-blur conditions, was expected to have two alternative, mutually-exclusive hypotheses. First, if external features act as a distraction during face selection, thereby providing no contextual benefit for selection of internal features, then naming should be positively related to degree of manipulation. Specifically, higher levels of blur should lead to better naming, and so low and medium levels would promote performance somewhere between no-blur and high-blur. Second, if external features do have some utility for face construction, then blurring would itself interfere to some extent with external retrieval cues used for face selection. So, while high blur should be more effective than no blur at all, less intense levels of blur would be even better. According to this second explanation, naming accuracy should be higher when images are produced under intermediate (low, medium or both) relative to high blurring.

### **Stage 1: Face Construction**

#### **Method.**

**Constructors.** Forty students were recruited to construct the composites by



opportunity sampling at the University of Central Lancashire. They did not participate in any other task in this paper (the same for participants involved in other experiments). There were 27 male and 13 female volunteers with an age range from 19 to 25 years ( $M = 21.0$  years,  $SD = 1.7$  years).

**Materials.** Target photographs were acquired from 10 male football players of the Comrades Sunday league team, Rotherham, UK; all were in their late-teenage years or early twenties. Images were captured in a front-facing pose with good lighting. Facial expressions were neutral. Each subject had minimal facial hair and none wore glasses. Photographs were cropped using Adobe Photoshop into a headshot that revealed shoulders, but left the background unaltered. They were printed in colour to dimensions of approximately 8cm wide by 10cm high. An example target photograph is shown in Figure 2.

**Design.** A design was chosen that reflected real-life composite construction in the laboratory as far as possible (e.g., Frowd, Carson, Ness, McIntyre, et al., 2005), to allow good generalisation of results. Target faces were unfamiliar to constructors, but familiar to (participant) evaluators who would later assess their quality by naming them. After having inspected a target face, each constructor waited 22 to 26 hours before undergoing a cognitive interview to recover a description of the face. The interview used was typical of the type administered in UK investigations for face construction (Frowd, Carson, Ness, Richardson, et al., 2005). This included rapport building, to help constructors relax; context reinstatement, to assist in visualising the face; exhaustive recall of the face in a ‘free’ format; and ‘cued’ recall, to prompt for further details. This was followed by face construction using EvoFIT. Excepting the blur setting, this system was used in the same way as it would be with real witnesses (Frowd, Hancock, et al., 2011). The role of the operator (experimenter) was to operate the software and to guide participant-constructors through the process. Initially, each constructor selected external features to match those of the target. Faces were then generated and presented in arrays for selection with these external features at one of four levels of blur (none, low, medium or high). For the final array of faces, the blur, if used, was disabled, and constructors selected a single best face, before attempting to improve its likeness using the holistic and shape tools.

Frowd, Pitchford, et al. (2010) reported a very large effect size on naming when blurring was used at face construction. Our design aimed to detect large effect sizes also at the construction stage. A 2x4 repeated-measures design was used, with the first factor being type of face presented at face naming (complete or internal features; see Stage 2, below) and the second factor being level of blur used at construction. Using a by-items composite naming analysis, a G\*Power computation (Faul, Erdfelder, Lang, & Buchner, 2007) predicted a minimum requirement of 8 targets ( $f = 0.4$ ,  $\alpha = .05$ ,  $1-\beta = .75$ ,  $r = .8$ ): we chose to exceed this requirement by using 10 targets.

***Procedure.*** Participant-constructors were tested individually throughout. Each person was randomly assigned to one of four blurring conditions (none, low, medium or high) within the constraint that each of the 10 target faces would be constructed once in each blurring condition—this design produced a total of 40 composites. A constructor was given a randomly-selected photograph to inspect for 60 seconds; this was carried out in the knowledge that a composite of the face would be constructed on the following day. The experimenter remained blind to the identity of each target until all composites had been constructed.

Between 22 and 26 hours later, each person returned to the laboratory to construct a composite. Sessions were self-paced and initiated by an informal dialog between experimenter and constructor. It was explained that a cognitive interview would be administered first, to help the person recall details of the target face, and that a composite of it would then be constructed using EvoFIT. Each person was asked to think back to the previous day and try to visualise the face that had been seen, and then to freely recall it in detail when they were ready. The experimenter also mentioned that she would not interrupt while the face was being described. After the constructor had given a description, the experimenter repeated it back, pausing at each individual feature to give the person the opportunity to attempt further recall.

The session moved on to composite construction with EvoFIT. The procedure used was detailed and is described in earlier work (Frowd, Bruce, Ness, et al., 2007; Frowd, Pitchford, et al., 2010); a summary is provided here. The experimenter presented the constructor with a selection of external features for selection of the best likeness. The blur level was set according to the assigned condition and, with the exception of the condition in which blurring was not used, the experimenter mentioned that the blur would help each person to select faces. Each constructor was presented with a screen displaying facial shapes and told that the images varied by feature shape and feature placement. The experimenter requested selection of six whole faces in total from four of these shape displays. Afterwards, facial textures were shown and the experimenter explained that the faces now varied by greyscale shading for eyes, brows, mouth and overall skin tone. Each person was again asked to select six items from four of these screens. The selected items were then bred together, to combine characteristics, and the selection procedure to acquire six shapes and six textures was repeated. Blurring, if used, was then disabled. The best-matching face judged by the constructor was imported into EvoFIT's Holistic tools for adjustment of age, masculinity and other whole-face properties. The face was further enhanced, if required, by changing size and placement of features using the Shape tool. The resulting image was saved. Sessions took approximately one hour to complete.

## **Stage 2: Composite Evaluation**

## **Method.**

**Evaluators.** Thirty participants volunteered as evaluators. All were from the Rotherham area in the UK, and followed the Sunday league football team from which the target photographs were taken. Complete composites were inspected by one group of 15 evaluators, 9 males and 6 females, whose ages ranged from 18 to 31 years ( $M = 23.8$  years,  $SD = 3.5$  years). A second group of 15 evaluators studied internal composite features; these were 6 males and 9 females, age ranging from 21 to 36 years ( $M = 23.7$  years,  $SD = 3.8$  years).

**Materials.** Complete composites were printed using a good-quality greyscale printer on single sheets of A4 to dimensions of approximately 6 cm wide by 8 cm high—printing was carried out in greyscale as EvoFIT synthesises faces in this image mode (note that evolving in colour does not appear to be advantageous relative to greyscale: Frowd, Bruce, Plenderleith, & Hancock, 2006). The Oval and Rectangular Marquee tools in Adobe Photoshop were employed to crop internal features from complete images. Examples are presented in Figure 2. A set of target photographs was printed, as in Stage 1.

Figure 2 about here

**Design.** The 40 printed composites produced in Stage 1 were given to participants who were familiar with the target identities for naming; recall that this set contains 10 composites constructed at each of the four levels of blurring. Two groups of participant-evaluators were recruited in Stage 2: the first group inspected complete composites, while the second were shown composites with external features removed, to verify that the blur manipulation affected the (important) internal-features region of the composites. The two alternative hypotheses for the experiment are detailed above.

A within-subjects design was used for blur type (none, low, medium or high), with participant-evaluators inspecting all 40 composites. Image type (complete or internal-features composites) was between-subjects.

**Procedure.** Participant-evaluators were tested individually. Each person was told that composites of footballers from the Comraides Sunday League football team would be seen and to name as many as possible. It was mentioned that the set contained repeated identities; also, that it was permissible to pass on to the next image without giving a name—a ‘don’t know’ response. The relevant set of printed composites (complete for one group and internal features for another) was presented sequentially, in a different random order for each person, with each evaluator offering a name as requested. Afterwards, a printed copy of the 10 target photographs used to construct the composites were presented, also in a different random order for each person, and evaluators similarly attempted to name these images (as a target

familiarity check). The task was self-paced.

## Results.

Naming of target photographs, carried out after naming of the composites, indicated that evaluators were, as expected, very familiar with the relevant identities ( $M = 90.7\%$ ,  $SD = 3.4\%$ ). The raw naming data for composites were scored for accuracy with respect to the relevant target; the means of these data (expressed in percent correct) by condition are detailed in Table 1. There was a trend such that composites were correctly named with increasing success as the level of blur increased from none to low to medium to high. Mean naming was also higher at each blur level when shown complete, relative to internal-features images, except for when blurring was not used (see column labelled 'None' in the table) where mean naming was identical.

Table 1 about here

Analysis of Variance (ANOVA) was conducted on the mean correct naming scores by-items rather than by-participants, as we wish to generalise across composites rather than across recognisers—it is also the case that the latter analysis tends to be sensitive to outlier effects, a particularly good- or bad-quality composite, which can skew the interpretation. Note that the by-items analysis here is repeated measures for both factors. In the ANOVA, Mauchly's test of Sphericity was not significant for image type or blur level (Mauchly's  $W > 0.7$ ) but was significant for the interaction term, and so the degrees of freedom for this factor were corrected using Greenhouse-Geisser ( $W = 0.16$ ,  $p = .014$ ).

ANOVA was significant for image type [ $F(1,9) = 38.4$ ,  $p < .001$ ,  $\eta_p^2 = .81$ ] as naming was higher for complete than for internal-features composites. Blur level also exerted a significant effect [ $F(3,27) = 137.0$ ,  $p < .001$ ,  $\eta_p^2 = .94$ ], and two-tailed repeated contrasts confirmed that naming reliably increased from no blur to low blur ( $p = .017$ , Cohen's  $d = 1.5$ ), from low to medium ( $p < .001$ ,  $d = 3.7$ ) and from medium to high ( $p = .044$ ,  $d = 0.9$ ). The interaction between these main effects was also significant [ $F(1.8,16.3) = 6.6$ ,  $p = .009$ ,  $\eta_p^2 = .42$ ]. This was partly due to floor effects, which are highlighted below in the Discussion. The interaction was also significant since intact (complete) composites promoted better naming than internal-features composites when blur was used (low, medium and high) ( $p < .05$ , all one-tailed) but not in the no-blur condition ( $p = 1.0$ ). Further, greater levels of blur yielded more correct naming ( $p < .005$ ) except when comparing no- with low-blur internal-features composites, and medium- with high-blur for complete composites: in both cases, there was no reliable change ( $p > .16$ ).

To provide a supplementary measure of composite quality, data were re-scored

(by-items) according to the number of incorrect names given by evaluators. This metric is valuable, since, *per se*, lower values indicate more accurate constructions; in an applied context, fewer incorrect names (e.g. given by members of the public) reduce the number of false leads that the police must follow-up. Here, incorrect naming was infrequent, with a mean of 6% or less in each cell of the design (for both complete and internal features images)<sup>1</sup>. No inferential statistics were conducted due to low mean values.

## Discussion.

Experiment 1 explored the role of external-feature information in face construction using EvoFIT. Participant-constructors inspected an unfamiliar face, and then one day later evolved a composite of it by selecting from face arrays with one of four levels of external-features blur. The resulting images were named by other people (participant-evaluators). These data indicate an overall pattern of effects: higher levels of blur promoted higher levels of correct naming. The significant interaction between blur level and face type also indicates that internal features contributed to naming of complete images under all levels of blur, except for constructions that did not use blur. It was in the no-blur condition that the influence of external features was largest, and task performance was worst. These results support Frowd, Pitchford, et al. (2010), who similarly found very low correct naming for intact images constructed with EvoFIT without blur ( $M < 5\%$ ).

The interaction term also indicates that two further contrasts failed to reach significance, despite trending in the direction of the main effect for blur: from no- to low-blur when naming internal features ( $d = 0.8$ ), and from medium- to high-blur for intact-image naming ( $d = 0.3$ ). In the former contrast, which should have had sufficient experimental power ( $1-\beta = .75$ ), the null effect arose due to a floor effect—but power was lacking in the latter contrast ( $1-\beta = .2$ ). In any case, it is the main (overall) effect of blur type that is of importance here.

The main-effects analysis demonstrated that intermediate levels of blur yielded intermediate-quality composites, favouring the hypothesis that external features operate as a distraction rather than as a beneficial retrieval cue for face construction. (If external features were a useful cue, one or both of the intermediate levels would have elicited better naming than that observed for high-blur composites.) As blurring and composite quality are thus positively related, it is possible that even more extreme levels of blur (greater than 8 cycles per f/w) may promote still better performance. This idea was tested in the following experiments.

---

<sup>1</sup> For clarity, naming data consists of scores for correct names ( $M = 38.8\%$ ), for incorrect names ( $M = 3.6\%$ ) and for cases where no name was offered ( $M = 57.6\%$ ).

## Experiment 2

Although Frowd and Hepton (2009) did not use external-features blurring, they did find that increasing the accuracy of hair shown in EvoFIT face arrays promoted better-quality composites. Specifically, internal features were constructed more identifiably when hair was a good match to the target's than when the match was poor. Composites were constructed best, though, when hair was presented exactly as seen in the target photograph. While this is an artificial situation, as "exact" hair is not available to police witnesses who construct composites, the effect size of the difference between similar and exact hair was large, indicating constructors' considerable sensitivity to this feature during face construction. Here, we attempted a partial replication of this effect by having face construction proceed using hair that was either similar or an exact match to the target photograph. We included a third condition, designed to provide even more accurate contextual information and thus improve face selection further, by presenting face arrays containing hair and shoulders from the target photograph. Our prediction was that composite naming—especially of the important internal features region—would improve from 'similar hair' to 'exact hair', replicating Frowd and Hepton, and from 'exact hair' to 'exact hair and shoulders'.

A fourth condition was included that did not present external features at all during construction—equivalent to infinite level of blur. If external features promote accurate selection from face arrays, then an alternative hypothesis is that composites constructed under the 'internal-features only' (internals-only) condition should be recognised worst, lacking external-feature cues. If, however, external features serve as a distraction, as suggested in Experiment 1 and elsewhere (Bruce et al., 1999; Frowd, Bruce, McIntyre, et al., 2007), then the second (and so more likely) alternative hypothesis is that the most identifiable composite faces should result from their absence.

### Stage 1: Face Construction

#### **Method.**

**Constructors.** Participants who constructed the composites were volunteer students from the University of Central Lancashire, 20 males and 20 females. Their age ranged from 19 to 24 years ( $M = 21.0$  years,  $SD = 1.4$  years).

**Materials.** Materials were target photographs of 10 male officers from Greater Manchester Police (GMP), Wigan, UK. Each person was photographed under good lighting, in a front-facing pose and with a neutral expression. The age of these photographic subjects ranged from 27 to 52 years ( $M = 40.8$  years,  $SD = 7.4$  years). Adobe Photoshop was used to crop the photographs and to copy hair and shoulders,

for importing into EvoFIT (for showing to constructors). ‘Similar’ hair for use in composite construction was chosen (by the first author) to be the closest visible match in the system. See Figure 3 for example external features. Photographs were printed in colour to portrait dimensions of approximately 5 cm wide by 8 cm high.

**Design.** The design and procedure were fundamentally the same as in Experiment 1, including a 20 – 28 hour delay to construction, a cognitive interview and face construction using EvoFIT. This time, blurring was not used. Instead, constructors produced a single face with external features either entirely absent from the face arrays, or which were shown in one of three ways: similar hair, exact hair or exact hair and shoulders—again, a between-subjects design. They did not select hair: it was selected for them as part of the experiment, appropriate for the particular target and condition to which they were (randomly) assigned.

**Procedure.** The procedure involved a cognitive interview and face construction with EvoFIT, as in Experiment 1, except that each evaluator here was randomly allocated, with equal sampling, to one of four external-features conditions (none, similar hair, exact hair or exact hair and shoulders), and no blur was applied to external features. Testing sessions lasted for approximately as long as before.

## **Stage 2: Composite Evaluation**

### **Method.**

**Evaluators.** Participants who volunteered for the naming exercise were staff from GMP, Wigan. Twelve (7 male and 5 female, with an age range from 26 to 48 years;  $M = 36.4$  years,  $SD = 6.5$  years) inspected intact (complete) composites and 12 (7 male and 5 female, with an age range from 28 to 47 years;  $M = 38.6$  years,  $SD = 5.5$  years) inspected composites with internal-features only.

**Materials, Design and Procedure.** Two groups of evaluators were again recruited to name printed composites seen in one of two image preparations. For both groups, evaluators were tested individually. The first group saw 30 complete composites produced under conditions in which external features were visible: similar hair, exact hair, and exact hair and shoulders. The second group saw these same images shown as internal features along with the 10 composites constructed in the internals-only condition, images that themselves contained just internal features. Afterwards, target photographs were presented for naming. Printing of composites and targets, their dimensions, and the procedure used to name the composites, were the same as in Experiment 1. Example images constructed in Stage 1 are presented in Figure 3.

Figure 3 about here

### **Results.**

Target photographs were again named very well ( $M = 88.3\%$ ,  $SD = 11.2\%$ ). The correct naming percentages for composites in each condition are presented in Table 2. When external features were present during construction, correct naming of complete composites increased from similar hair to exact hair to exact hair and shoulders. Naming of internal-features composites followed the same pattern of means, but was best when faces were constructed without visible external features.

Table 2 about here

The first analysis considered the impact of seeing external features—similar hair, exact hair, or exact hair and shoulders—when building the face, and so analysed the naming data for complete images (top row of Table 2). A repeated-measures ANOVA was significant for external-features type [ $F(2,18) = 40.3$ ,  $p < .001$ ,  $W = 0.58$ ,  $\eta_p^2 = .82$ ], and repeated contrasts indicated that exact hair elicited superior naming to similar hair ( $p = .002$ ,  $d = 1.9$ , one-tailed), and that exact hair and shoulders was likewise superior to exact hair ( $p < .001$ ,  $d = 2.0$ , two-tailed). For analysis of internal features, all four construction conditions were included (bottom row of table). The ANOVA was also reliable [ $F(3,27) = 89.3$ ,  $p < .001$ ,  $W = 0.41$ ,  $\eta_p^2 = .91$ ] as were all repeated contrasts: exact hair led to better naming than similar hair ( $p = .012$ ,  $d = 1.2$ , one-tailed), exact hair and shoulders led to better naming than exact hair ( $p < .001$ ,  $d = 2.8$ , two-tailed), and internals-only led to better naming than exact hair and shoulders ( $p = .001$ ,  $d = 0.9$ , two-tailed).

Next, we estimated the contribution to naming of external features themselves. To do this, mean composite naming data, by-items, were subtracted between complete and internal-features conditions (that is, data columns one to three in the table;  $MD = 40.0\%$ ,  $SD = 5.0\%$ ) and subjected to ANOVA. The analysis revealed that external features were important [ $F(2,18) = 4.8$ ,  $p = .021$ ,  $W = 0.9$ ,  $\eta_p^2 = .35$ ], with both exact hair ( $d = 1.2$ ) and exact hair and shoulders ( $d = 0.8$ ) yielding better naming performance than images seen with similar hair ( $p < .05$ , both one-tailed); no reliable difference emerged between faces composed in the presence of exact hair versus those composed with exact hair and shoulders [ $t(9) = 1.4$ ,  $p = .19$ , two-tailed].

Incorrect naming was infrequent, as in the first experiment; here, mean incorrect naming was less than 2% in each cell.

## Discussion.

In this experiment, we increased the accuracy and extent of external features in the arrays presented at face construction, successively increasing their correspondence with the target photograph. Correct naming of both complete and internal-features composites increased reliably with the successive addition of similar hair, exact hair, and exact hair and shoulders. Therefore, increasing the correspondence of the exterior



region promotes better quality composites, a finding that replicates and extends Frowd and Hepton (2009) using a set of different target identities.

It is not really surprising that correct naming of complete images increased over the three conditions, as closer-matching external features are themselves likely to be valuable to naming to some extent, but effect sizes for naming of internal-features composites were intriguing and so are worthy of further comment. The benefit of exact hair over similar hair was large ( $d = 1.2$ ), similar to the advantage found in Frowd and Hepton (2009). However, relative to exact hair, there was a huge additional advantage for construction in the presence of exact shoulders ( $d = 2.8$ ). Note that this advantage relates to face construction, since these data are for naming of internal features only. It is likely that providing some body context allows a constructor to achieve a better sense of the overall appearance of the face at each stage in the process: in essence, that it assists with whole-face selection and manipulation. This may also explain, in one of our early EvoFIT interfaces, why constructors could evolve identifiable likenesses with fairly good (unblurred) matching hair and shoulders (Frowd, Bruce, Plenderleith, et al., 2006). Future work could now systematically explore the impact of shoulders (and even additional body information) as a likeness to the target rather than as an exact match, which is the normal case. The large benefit conferred by shoulders is considered further in the General Discussion.

Naming of the internal-features composites reveals a second important finding: construction without any external features (internals-only) promoted much more identifiable internal features than construction with exact hair and shoulders. While the latter condition provides an exterior context that closely reflects what the constructor saw, that region of the image will still contain inconsistencies. First, some exterior parts do not match, in particular the background context—notice, for example, the textual background (a door) behind the subject in Figure 2(e). Such discrepancies may give rise to distraction, worsening performance. Second, as a result of using evolving systems to produce facial shapes, the entire exterior region is distorted—this includes hair, forehead, ears, neck and (if present) shoulders. Generation of shape information in this way leads to visible differences in this region with respect to the target, and consequently, potential distraction. In this experiment, external features were pre-selected: should constructors have selected them for themselves from a range of alternatives, discrepancies would have been larger, and performance worse.

So, for face construction, our results lend support to the suggestion that external features function as a distraction, that the amount of distraction they exert is inversely related to the accuracy and extent of their match with a target (similar hair < exact hair < exact hair and shoulders), and that mismatches can be limited to some extent (using exact hair and shoulders), but that task performance is superior when external

features are not visible at all during construction. These results, taken together with results of Experiment 1, suggest that evolving without external features should promote a more identifiable image than evolving using a high level of external-features blurring, the current police practice for using EvoFIT. This hypothesis was tested in the following experiment.

### **Experiment 3**

We compared composites constructed using the standard EvoFIT procedure, with high blur applied to external features during face selection (what we refer to here as ‘standard-blur’) against composites constructed without visible external features (‘internals-only’). In the previous experiment, for construction with internals-only, hair and other external features were not added; in this experiment, constructors selected them after reproducing the internal features. While it is well established that naming familiar faces is largely dependent on internal features, exterior features do contribute to some extent for both photographs of faces (e.g. Ellis et al., 1979) and for composite faces (e.g. Frowd & Hepton, 2009; Experiment 2 here). Therefore, their presence is of value for composite recognition (e.g. by members of the public). In the internals-only condition, once external features had been added, constructors were offered the opportunity to improve the likeness by manipulating feature shapes and holistic properties of the face: they could, for instance, alter the hairline, the positioning of internal features, and the shape and position of ears.

In our other construction condition, standard-blur, blurring was removed after constructors had completed face selection (in the same way as it was in Experiment 1 for conditions using blur). We hypothesised that, in this condition, external features would now exert a strong negative influence on the remaining construction process—that is, external features would be seen clearly during adjustment of holistic and shape aspects of the face. It was anticipated that any such negative influence of external features would yield worse likenesses than if the exterior context remained blurred. To test for this possibility, we included a third condition that maintained blurring until the internal features had been constructed in their entirety (‘extended-blur’). This condition maintained external-features blurring on the best face while constructors made adjustments to it using holistic and shape tools. The hypothesis was that extended-blur would promote more identifiable composites than would standard-blur. As with evolving using internals-only, once blur was removed, constructors were given the opportunity to enhance shape and holistic properties. Note that extended-blur and internals-only were very similar conditions, differing only in what the constructor saw in the exterior region whilst reproducing the internal features: blurry external features (extended-blur) and uniform light-grey texture (internals-only).

We predicted that extended-blur would promote better-quality composites than

standard-blur, but that best overall performance would emerge from face construction in the internals-only condition, due to the absence of distracting external features.

## **Stage 1: Face Construction**

### **Method.**

**Constructors.** Participants were each paid £5 to construct a composite. They were recruited from staff at two companies in NW England: Taylor Wimpey, Wakefield; and Eliot Gardens, Wigan. Constructors were 15 males and 15 females with an age range from 25 to 55 years ( $M = 36.3$  years,  $SD = 8.4$  years).

**Materials.** Targets were 10 front-facing, good-quality colour photographs of international-level UK footballers, sourced from the Internet; these identities are listed in Table 3. The photographs were selected on the basis of near-neutral expression, not wearing glasses and being clean-shaven or with only minor stubble. Cropping maintained shoulders as in Experiment 1. They were reproduced in colour as before.

**Design and Procedure.** The basic design and procedure was again the same as in Experiment 1. This time, individual constructors were randomly assigned to construct a single face using standard-blur, extended-blur or internals-only. Each person inspected a target face, randomly selected, after verifying that the face was unfamiliar to them (had anyone recognised their given face, another would have been randomly selected). Each person then described and constructed a composite of it after 22 to 26 hours. For standard-blur, construction followed the same procedure as used in Experiment 1: each constructor initially selected a set of external features, which were then presented highly blurred in the face arrays until completion of face selection. This same procedure was also followed for composites produced using extended-blur, except that the experimenter did not disable external-feature blurring until each constructor reported that the best likeness had been achieved; thereafter, blur was deactivated and shape/holistic tool use offered again (i.e. it was offered for the first time prior to the blur being turned off). In the internals-only condition, external features were not selected until the best likeness had been reached; after adding the external features, shape/holistic tool use was again offered (also for the second time). Example likenesses produced are presented in Figure 4.

## **Stage 2: Face Evaluation**

### **Method.**

**Evaluators.** Forty-eight volunteers were recruited on the basis of being followers of UK international-level football. They were players from the Wigan hockey team, Wigan, and staff from Taylor Wimpey, Wakefield. There were 44 males and 4 females and their age ranged from 18 to 60 years ( $M = 28.5$  years,  $SD = 11.7$  years).

**Materials.** In the previous naming paradigms, participants attempting to recognise the composites knew the identities well (i.e. target naming was about 90% correct). For naming of composites from police investigations, members of the public are unlikely to have such familiarity with offenders. So, to make the current naming task more realistic, four ‘filler’ items were included in the composite set shown to participant-evaluators. This reduces evaluators’ ability to make an identification by process of elimination from the somewhat limited pool of international footballers: they know that some composites are not of footballers and so have to make a more active identification. These filler items were randomly generated from EvoFIT; they were faces selected (by the first author) from the first generation to appear visually similar to the footballer targets. All were given a short, contemporary hairstyle typically worn by these athletes. Composites and target photographs were printed as previously.

Figure 4 about here

**Design and Procedure.** In the two experiments described previously, evaluators inspected all of the composites produced during Stage 1. This within-subjects design has good experimental power, but repeating identities while presenting items can inflate naming levels as a result of inter-item cueing—that is, a good quality composite can act as a prompt for another composite of the same identity. In the current experiment, we adopted the more realistic stance of naming *without* repeated identities; also, as described in the Materials, the realistic element was extended to include four additional filler composites that were not of footballer targets.

We implemented a between-subjects design: each evaluator was randomly allocated to one construction condition (normal-blur, extended-blur or internals-only), in which they were asked to name a set of composites. Each person also saw the same four filler composites mixed in randomly among the true composites. At the start, the experimenter explained that some of the composites to-be-seen were based on UK international football players, and the task was to name as many of these identities as possible. Otherwise, the procedure for composite and target naming was the same as that described previously.

## **Results.**

Evaluators were very familiar with the target photographs: correct naming by condition ranged from 90.6% to 93.8%. The design of this composite naming study is between-subjects for condition (normal blur, extended blur and internals-only); it differs from the above, which were within-subjects (note that this is not to be confused with the *analysis* of naming data, which is within-subjects by-items in all experiments). For the current design, we must therefore be careful to guard against

chance sampling differences in target naming influencing composite naming, and so so used a different measure here, ‘conditional’ naming—as advised by Frowd, Carson, Ness, Richardson, et al. (2005). So, after calculating naming accuracy scores, conditional scores were calculated by dividing the number of times a (non-filler) composite was named correctly by the number of times its respective target was named correctly. Table 3 shows these adjusted mean scores as percentages. Internals-only construction recovered the most identifiable faces, followed by extended-blur and then normal-blur.

Table 3 about here

A one-factor repeated-measures ANOVA revealed a significant effect of external-features type on conditional naming scores [ $F(2,18) = 5.6, p = .013, W > 0.9, \eta_p^2 = .38$ ]. One-tailed simple contrasts revealed that both extended-blur ( $p = .049, d = 0.7$ ) and internal-only ( $p = .003, d = 1.3$ ) conditions produced superior composites to normal-blur. A one-tailed paired-samples  $t$  test comparing extended-blur and internals-only was not significant [ $t(9) = 1.7, p = .12$ ].

The percentage of composites named incorrectly was similar for normal-blur ( $M = 28.1\%, SD = 4.4\%$ ) and for extended-blur ( $M = 24.4\%, SD = 17.0\%$ ), but both of these conditions yielded somewhat higher levels of misnaming than when composites were constructed with internals-only ( $M = 17.5\%, SD = 10.5\%$ ); recall that lower levels of incorrect naming indicate more accurate likenesses. ANOVA produced a result that was not significant [ $F(2,18) = 3.3, p = .060, W = 0.52, \eta_p^2 = .27$ ]. Simple contrasts (planned, two-tailed) revealed no significant difference between normal- and extended-blur ( $p = .51$ ), but significantly fewer incorrect names were produced for internals-only than for normal-blur construction ( $p = .012, d = 1.2$ ). (Inferential statistics were not run between extended-blur and internals-only given that the ANOVA had yielded a non-significant result.)

We evaluated names offered for filler composites across conditions (or booklets). Recall that the same four unfamiliar faces appeared in each booklet; any name given for these items is, by definition, incorrect. Overall mean foil naming was 34.9%, but the range of means by condition was very small ( $MD = 1.6\%$ ) [ $F < 0.1$ ], indicating that evaluators’ willingness to offer any name varied little by construction condition.

## Discussion.

In this experiment, we followed-up on the results of our first two experiments, which suggested that constructing composites without any external features produces more identifiable faces than when external features are presented with high blur during face selection. That work was extended here, with the internals-only condition leading to higher correct and lower incorrect naming than in the normal-blur

condition. This study also lent support to our suggestion that prolonging blurring until later in the construction process would further reduce the negative impact of external features and promote more identifiable likenesses, relative to when blur was disabled mid-session. So, while the internals-only method did not promote reliably more correct composite naming than for extended-blur, this condition did result in significantly fewer incorrect names from its composites, suggesting overall superiority for construction by internals-only.

## **General Discussion**

Recovering facial memories is of theoretical interest in understanding how we access stored memories, but is also, if done accurately, of forensic value to police investigations. This paper focused on the role played by external features when constructing faces using a modern ‘evolving’ system, EvoFIT, and sought to resolve a ‘tension’ between two competing explanations—namely that, during face construction, external features constitute either a useful retrieval cue, or a distraction.

The first experiment demonstrated that progressively decreasing the visual presence of the external features by increasing the Gaussian blur applied to them promoted construction of progressively better-named images. Therefore, external features do not appear to have a useful function for face construction with EvoFIT. Experiment 2 varied the accuracy (similar hair to exact hair) and extent (exact hair to exact hair-and-shoulders) of external features available to composite constructors, and found that composite naming increased correspondingly, but not as much as when external features were masked throughout construction. This result further highlights the negative role played by external features: reducing their mismatch with a target face yields better-quality composites; also that, even when exactly matching the target, external features still interfere since masking them from the arrays improves composite quality (relative to the other conditions when external features were seen). This experiment also demonstrated that presenting hair exactly as seen in the photograph helps evaluators to name the relevant identity, but there was no additional gain when exact shoulders were also shown. In Experiment 3, masking external features until the end of face construction led to more identifiable composites relative to the previous method of blurring external features during the initial face-selection stage; we also found an advantage for extending the ‘high’ level of blur until the final stage of the process, indicating that external features also negatively impact use of holistic (ageing, weight, health, etc.) and feature-manipulation tools.

Overall, then, the contribution of the work is a cohesive argument demonstrating the negative influence of external features for face construction with EvoFIT, and a better way to evolve a face—by constructing internal features, in their entirety, in the absence of external features.

## Theoretical Account

Unfamiliar face recognition is facilitated by the exterior facial context (e.g. Cutler et al., 1986; Ellis et al., 1979; Memon & Bruce, 1983; Sporer, 1993). Aspects of a probe face that overlap with those seen at encoding provide retrieval cues that serve to improve the accuracy of unfamiliar face recognition—a paradigm sometimes referred to as encoding specificity (Tulving & Thomson, 1973). Examples include head pose, facial expression, background context, clothing and hair. Face recognition is a holistic process, in which individual features and relational properties of the face influence each other (e.g. Davies & Christie, 1982; Tanaka & Farah, 1993).

In an attempt to produce a superior interface to human memory, modern methods of face construction are modelled on important components of recognition processes. Constructors are now presented with arrays of complete faces, with the properties of each face matching those of the target—for example, faces appear within the appropriate age range (for EvoFIT, within  $\pm 5$  years). The task involves repeatedly selecting whole faces from arrays, with ‘breeding’. The resulting face is assumed to have attained a likeness sufficient for enhancement by holistic-type manipulations such as changing perceived age, weight and masculinity (Frowd, Bruce, McIntyre, et al., 2006)—but changes to feature shape, position or both are also possible, although such requests tend not to alter the face greatly (Frowd, Pitchford, et al., 2010).

Since the internal features of a face are the most valuable for recognition (e.g. Ellis et al., 1979; Frowd, Bruce, McIntyre, et al., 2007; Frowd, Skelton, et al., 2011), EvoFIT models the construction process further by blurring external features during face selection, with the aim of promoting better construction of internal features. The results of research here and elsewhere (Frowd, Park, et al., 2008; Frowd, Pitchford, et al., 2010) have shown that this aim was vindicated: our current work demonstrates that, far from being a useful retrieval cue, external features are a distraction. How, then, do we square our findings for face construction with established research that external features are a *useful* retrieval cue for face recognition?

It is worth noting that face construction is different to face recognition. While both tasks visually match facial information to memory, the former involves imperfect representations (e.g. Frowd, Bruce, Ross, McIntyre, & Hancock, 2007). With composite construction, faces are selected that have differing degrees of error to a target; to complicate this issue, different parts of the face are likely to be inaccurate by differing degrees and, for EvoFIT, between one face and another in the presented arrays. To complicate matters further, individual features and their placement on the face provide contextual information that interact with each other (e.g. Davies & Thomson, 1988). Increasing the distance between eyes and mouth, for example, elongates the appearance of a face (Yasuda, 2005): long, bushy hairstyles make a face appear wider than do short styles.

Changing between similar hairstyles in face arrays is unlikely to fundamentally change the overall perception of a face, as opposed to when shoulders are added. In preliminary work, the perceptual effect of concealing shoulders was markedly different for different participants, increasing or decreasing face width—readers are invited to try this for themselves by concealing shoulders in Figure 3d. As such, presence of shoulders is likely to reduce perceptual error and thereby increase accuracy of selection in face arrays. The impact of this was seen in Experiment 2: naming of the internal features of the composite vastly increased ( $d = 2.8$ ) when exact hair-and-shoulders were shown in the arrays rather than when just exact hair was shown. The experiment also found that naming was not facilitated by presenting images with shoulders present (relative to when exact hair was present), and therefore their impact is relevant to face construction and not to face naming. Other research using non-composite media has found that whole-body information—video footage including shoulders—only slightly modulates recognition of a familiar person (e.g. Burton et al., 1999).

We argue then that concealing shoulders has a detrimental effect at construction, but other factors are at play. Experiment 2 revealed that even with arrays containing highly-accurate external features (exactly-matching hair, ears and shoulders), naming of the resulting composites improved when all external features were omitted in the arrays. Similarly, the same experiment, along with Frowd and Hepton (2009), found that naming markedly improved ( $d = 1.2$ ) following only minor changes to hair (from similar to exactly matching in face arrays). So, the mere presence of external features subsequently reduces the identifiability of a composite's constructed internal features.

One plausible explanation is that when external features are visible at initial encoding, users will later attempt to construct them, at the cost of paying attention to internal features: when external features are present, there is simply more information for a constructor to process—in effect, the search space reduces in size when external features are not seen. Reduction of the search-space also occurs, although to a lesser extent, when external features are blurred, and also when this blurring persists until the end of face construction. There is some evidence to support this idea: in Frowd, Park, et al. (2008), constructors were presented with faces that had either a sketch-like or photo-quality appearance. While blurring was used in both conditions, more recognisable images were produced using EvoFIT from the less complex representation. It is also possible that composites produced by a sketch artist, which also tend to contain less detail than photographic composites, are for the same reason more recognisable than those made using feature systems after long delays (Frowd, Carson, Ness, McIntyre, et al., 2005; although, see also Davies & Little, 1990).

The advantage reported here of masking external features may also extend to different face-production systems. The traditional method is to select individual facial features to compose a single face. When doing so, there is a tendency for constructors



to first request selection of hair and face shape (Frowd, Carson, Ness, Richardson, et al., 2005)—two cardinal features of unfamiliar face perception (Ellis, 1986). When using traditional methods of this kind, we would predict that better construction of internal features is more likely if external features are masked (or blurred) at the start, and then selected after finalising the internals. Technology of this type tends to yield poor likenesses if there is more than a few hours delay between seeing the target and making the composite (e.g. Frowd, Carson, Ness, McIntyre, et al., 2005; Frowd, Bruce, Ness, et al., 2007; Frowd, Pitchford, et al., 2010), and so any means of facilitating face construction would be worthwhile. Ongoing research is exploring this issue.

### **Value of External Features for Composite Naming**

Frowd and Hepton (2009) estimated the contribution to naming of external features by subtracting naming scores from complete and internal-features-only images (as was done here in Experiment 2). Calculated in this way, external features here made the same contribution to the correct naming of composites irrespective of whether the composite was seen with similar or exactly-matching hair (Mean Difference,  $MD = 16\%$ ). The contribution of external features remained similar ( $MD = 9\%$ ) even when internal-features naming was very poor ( $M = 3\%$ ), suggesting that the extent to which external features contribute to recognition is both consistent and limited. Frowd, Skelton, et al. (2011) used a more direct measure, by presenting participant-evaluators with actual external features of composites, and found naming to be very similar ( $M = 13.2\%$ ) to Frowd and Hepton; they also found that naming of external features did not change reliably under different methods of face construction (feature or evolving), underscoring the importance of internal features for composite naming.

Calculated in the same way for our first experiment, external features have some contribution to naming ( $MD = 8.5\%$ ), consistent with the above estimates. In Experiment 2, this figure was much higher ( $MD = 40.0\%$ ), which we believe to be a consequence of the wider target age range (27 – 52 years) than that used in Experiment 1 (all < 30 years). Simply put, more variable age cues in the external images from Experiment 2—for example, grey, greying or non-grey hair colour, extent of forehead and neck wrinkles—can be more useful for assessing identity than cues available from younger targets (Frowd and Hepton, 2009; Frowd, Skelton, et al., 2011). Also calculated using this ‘mean difference’ method, while Frowd and Hepton (2009) found no difference in naming when composites’ external features included similar and exact hair to the target, the results presented here found exact hair to be a better cue than similar hair, with a large effect size ( $d = 1.2$ ). This is likely to be caused by the same mechanism: our results suggest that composites of older faces provide more valuable cues as to the target’s identity. Lastly, we found no reliable difference in naming when composites’ external features included exact hair than

when they were presented with exact hair and shoulders. This suggests that any additional cues provided by inclusion of exact shoulders do not present additional value for identification. It may be that people's attire in general is not particularly useful for recognition, or simply that the clothing worn by the targets used here was not sufficiently distinctive to be of value. Future work in this area might usefully provide some pertinent insight, since in the forensic setting, composites are occasionally produced with shoulders included.

### **Generalisation of Results**

Composite naming scores in Experiment 1 and 2 were fairly high, with overall means in the region of 40% for complete images. This performance level is the likely result of inter-item cueing, since evaluators inspected multiple composite images of the same target identity. However, there were no such repeats in Experiment 3; this experiment also presented filler (unfamiliar target) faces, with the aim of promoting more realistic assessment of composite quality. Composites made under normal-blur were named correctly 22.7% of the time, a success-rate consistent with previous research using similar face-production and -evaluation methods (Frowd, Nelson, et al., 2011: 24.1%; Frowd, Park, et al., 2008, Study 3: 19.8%; Frowd, Pitchford, et al., 2010: 24.5%). Incorrect naming levels are overall comparable (here: 28.1%; *ibidem*: from 15 to 50%).

These comparable correct-naming results also hint that mode of target presentation is not an important factor—these studies used static photographs (Experiment 3 here; Frowd, Park, et al., 2008; Frowd, Pitchford, et al., 2010) and videos (Frowd, Nelson, et al., 2011). The same conclusion has been reported in a meta-analysis for face construction using Identikit and Photofit systems (Meissner & Brigham, 2001). Similarly, type of target familiarity does not seem to affect composite naming either—they were celebrities (footballers and snooker players) in both Experiment 3 and Frowd, Pitchford, et al. (2010), and personally-familiar (shop assistants and members of a university department) in both Frowd, Nelson, et al. (2011) and Frowd, Park, et al. (2008). This result also applies to feature systems, as evidenced by similar naming levels from studies that have used a short delay-to-construction and a modern feature system, but differ by familiarity type (e.g. Brace et al., 2000; Bruce et al., 2002; Davies et al., 2000; Frowd, Bruce, McIntyre, et al., 2007; Frowd, Carson, Ness, Richardson, et al., 2005).

Relative to normal-blur, the composites constructed in the extended-blur and internals-only conditions reported in Experiment 3 elicited large and very large effects, respectively; either method would therefore appear to be of practical value if used with evolving systems in criminal investigations. Of these two techniques, internals-only is arguably preferable since it appreciably reduces incorrect naming and consequently the number of false leads needing to be followed-up by police. The

internals-only method is currently being field trialled by police forces using EvoFIT in the UK, the US and mainland Europe. Results of the previous field trial indicate that composites can be constructed from EvoFIT which directly lead to the arrest of an offender in 38.5% of criminal investigations (Frowd, Hancock, et al., 2011).

## **Summary**

Face construction is used by police forces with the aim of detecting people who commit crime. Until recently, performance of face-production systems was poor, principally as the traditional feature-by-feature method used to construct the face is at variance with the whole-face method we naturally use to recognise faces. Recent systems for constructing facial memories now present witnesses and victims of crime with arrays of complete faces for selection, with cycles of breeding, to evolve a face. Here, we assessed the role of external features for producing an identifiable image using one of these newer systems, EvoFIT. It was found that external features are a distracting influence to the person evolving a composite, and that the best procedure is for these exterior features to be omitted until the rest of the face has been constructed. The approach should provide a marked improvement to correct naming of composites for police forces who use this type of technology, thereby promoting much better identification of police suspects; there is also the potential that the technique will generalise to more traditional methods of face construction (esp. feature and sketch systems), similarly benefitting police forces who continue to use older technology.

## References

- Andrews, T.J., Davies-Thompson, J., Kingstone, A., & Young, A.W. (2010). Internal and external features of the face are represented holistically in face-selective regions of visual cortex. *Journal of Neuroscience*, 30, 3544–3552. doi:10.1523/jneurosci.4863-09.2010
- Berman, G.L., & Cutler, B.L. (1998). The influence of processing instructions at encoding and retrieval on face recognition accuracy. *Psychology, Crime & Law*, 4, 89-106. doi:10.1080/10683169808401751
- 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.
- Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, 73, 105-116. doi:10.1111/j.2044-8295.1982.tb01795.x
- Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception*, 15, 387-397. doi:10.1068/P150387
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P.J.B., Burton, A.M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5, 339-360. doi:10.1037/1076-898X.5.4.339
- 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. doi:10.1037/0021-9010.87.5.894
- Bruck, M., Cavanagh, P., & Ceci, S. J. (1991). Fortysomething: Recognizing faces at ones 25th reunion. *Memory & Cognition*, 19, 221–228. doi:10.3758/BF03211146
- Burton, A.M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor quality video: evidence from security surveillance. *Psychological Science*, 10, 243-248. doi:10.1111/1467-9280.00144
- Campbell, R., Coleman, M., Walker, J., Benson, P. J., Wallace, S., Michelotti, J., & Baron-Cohen, S. (1999). When does the inner-face advantage in familiar face recognition arise and why? *Visual Cognition*, 6, 197-216. doi:10.1080/713756807
- Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception and Psychophysics*, 58, 602–612. doi:10.3758/BF03213093
- Craw, I., & Cameron, P. (1991). Parameterising images for recognition and reconstruction. *Proceedings of the British Machine Vision Conference*, Proceedings of the British Machine Vision Conference (pp-367-370). Springer Verlag: London.
- Cutler, B.L., Penrod, S.D., & Martens, T.K. (1987). Improving the reliability of

- eyewitness identifications: Putting context into context. *Journal of Applied Psychology*, 72, 629-637. doi:10.1037/0021-9010.72.4.629
- Cutler, B.L., Penrod, S.D., O'Rourke, T.E., & Martens, T.K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. *International Journal of Applied Social Psychology*, 2, 113-134.
- Davies, G.M., & Christie, D. (1982). Face recall: an examination of some factors limiting composite production accuracy. *Journal of Applied Psychology*, 67, 103-109. doi:10.1037/0021-9010.67.1.103
- Davies, G.M., & Little, M. (1990). Drawing on memory: Exploring the expertise of a police artist. *Medical Science and the Law*, 30, 345-354.
- Davies, G.M., & Milne, A. (1982). Recognizing faces in and out of context. *Current Psychological Research*, 2, 235-246. doi:10.1007/BF02684516
- Davies, G.M., Shepherd, J., & Ellis, H. (1978). Remembering faces: acknowledging our limitations. *Journal of Forensic Science*, 18, 19-24. doi:10.1016/S0015-7368(78)71178-3
- Davies, G.M., & Thomson, D.M. (1988). *Memory in context: context in memory*. London: Wiley.
- Davies, G.M., van der Willik, P., & Morrison, L.J. (2000). Facial Composite Production: A Comparison of Mechanical and Computer-Driven Systems. *Journal of Applied Psychology*, 85, 119-124. doi:10.1037//0021-9010.85.1.119
- de Haan, E.H.F., & Hay, D. (1986). The matching of famous and unknown faces, given either the internal or external features: A study on patients with unilateral brain lesions. In H.D. Ellis, M.A. Jeeves, F. Newcombe, & A. Young (Eds), *Aspects of face processing* (pp. 302-309). Dordrecht: Martinus Nijhoff.
- Ellis, H.D. (1986). Face recall: A psychological perspective. *Human Learning*, 5, 1-8.
- Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, 8, 431-439. doi:10.1068/p080431
- 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.
- Endo, M., Takahashi, K., & Maruyama, K. (1984). Effects of observer's attitude on the familiarity of faces: using the difference in cue value between central and peripheral facial elements as an index of familiarity. *Tohoku Psychologica Folia*, 43, 23-34.
- Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical Sciences. *Behavior Research Methods*, 39, 175-191.
- Frowd, C.D. (in press). Facial recall and computer composites. In C. Wilkinson and C. Rynn (Eds.) *Facial identification*.

- 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. doi:10.1348/000712606x104481
- 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. doi:10.4304/JMM.1.3.42-51
- Frowd, C.D., Bruce, V., Ness, H., Bowie, L., Thomson-Bogner, C., Paterson, J., McIntyre, A., & Hancock, P.J.B. (2007). Parallel approaches to composite production. *Ergonomics*, 50, 562-585. doi:10.1080/00140130601154855
- Frowd, C.D., Bruce, V., Plenderleith, Y., & Hancock, P.J.B. (2006). Improving target identification using pairs of composite faces constructed by the same person. *IEE Conference on Crime and Security* (pp. 386-395). London:IET. doi:10.1049/ic:20060341
- 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. doi:10.1080/13506280601058951.
- Frowd, C.D., Bruce, V., Smith, A., & Hancock, P.J.B. (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied*, 14, 276-287. doi:10.1037/1076-898X.14.3.276
- Frowd, C.D., Carson, D., Ness, H., McQuiston, D., Richardson, J., Baldwin, H., & Hancock, P.J.B. (2005). Contemporary Composite Techniques: the impact of a forensically-relevant target delay. *Legal & Criminological Psychology*, 10, 63-81. doi:10.1348/135532504X15358
- Frowd, C.D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S., & Hancock, P.J.B. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law*, 11, 33-52. doi:10.1080/10683160310001634313
- Frowd, C.D., & Fields, S. (in press). Verbal overshadowing interference with facial composite production. *British Journal of Psychology*.
- Frowd, C.D., Hancock, P.J.B., Bruce, V., Skelton, F.C., Atherton, C., Nelson, L., et al. (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. doi:10.1145/1008722.1008725
- Frowd, C.D., & Hepton, G. (2009). The benefit of hair for the construction of facial composite images. *British Journal of Forensic Practice*, 11, 15-25.
- Frowd, C.D., Nelson, L., Skelton F.C., Noyce, R., Heard, P., Henry, J., Morgan, D.,

- Fields, S., McIntyre, A., & Hancock, P.J.B. (2011). *Interviewing techniques for Darwinian facial composite systems*. Manuscript submitted for publication.
- 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.H., & Hancock, P.J.B. (2010). The psychology of face construction: giving evolution a helping hand. *Applied Cognitive Psychology*. doi: 10.1002/acp.1662
- Frowd, C.D., Skelton F.C., Butt, N., Hassan, A., & Fields, S. (2011). *Familiarity effects in the construction of facial composite images using software systems*. Manuscript submitted for publication.
- 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. doi:10.1504/IJESDF.2009.024900
- Hancock, P.J.B. (2000). Evolving faces from principal components. *Behavior Research Methods, Instruments and Computers*, 32-2, 327-333. doi:10.3758/BF03207802
- Hancock, P.J.B., Bruce, V., & Burton, A.M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4-9, 330-337. doi:10.1016/S1364-6613(00)01519-9
- Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 986-1004. doi:10.1037//0096-1523.22.4.986
- Hole, G.J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, 23, 65-74.
- Hole, G.J., George, P.A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, 31, 1221-1240. doi:10.1068/p3252
- Homa, D. Haver, B., & Schwartz, T. (1976). Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. *Memory & Cognition*, 4, 176-185. doi:10.3758/BF03213162
- Jeong, T., Kim, Y., & Lee, C. (2010). No-reference image-quality metric based on blur radius and visual blockiness. *Optical Engineering*, 49, 45001-45009.
- Johnston, R.A., & Edmonds, A.J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, 17, 577-596. doi:10.1080/09658210902976969
- Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive*

- Psychology*, 15, 101-116. doi:10.1002/1099-0720(200101/02)15:1<101::AID-ACP697>3.0.CO;2-7
- Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory & Cognition*, 27, 974-985.
- Leder, H., & Carbon, C.C. (2006). Face-specific configural processing of relational Information. *British Journal of Psychology*, 97, 19-29.  
doi:10.1348/000712605X54794
- Lee, K.J., & Perrett, D.I. (2000). Manipulation of colour and shape information and its consequence upon recognition and best-likeness judgments. *Perception*, 29, 1291-1312. doi:10.1068/p2792
- Lewis, M.B. (2004). Face-space-R: towards a unified account of face recognition. *Visual Cognition*, 11, 29-69. doi:10.1080/13506280344000194
- Macrae, C.N., & Lewis, H.L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science*, 13, 194-196. doi:10.1111/1467-9280.00436
- Meissner, C.A. & Brigham, J.C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, 15, 603-616.  
doi:10.1002/acp.728
- Meissner, C.A., Sporer, S.L., & Susa, K.J. (2008). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *European Journal of Cognitive Psychology*, 20, 414-455.  
doi:10.1080/09541440701728581
- Memon, A., & Bruce, V. (1983). The effects of encoding strategy and context change on face recognition. *Human Learning*, 2, 313-326.
- Mitchell, M. (1996). *An introduction to genetic algorithms*. MIT.
- Shapiro, P. N., & Penrod, S.D. (1986). Meta-analysis of facial identification rates. *Psychological Bulletin*, 100, 139-156. doi:10.1037//0033-2909.100.2.139
- Skelton, F.C., Frowd, C.D., & Speers, K.E. (2011). Context effects in the construction of facial composite images. Manuscript submitted for publication.
- Sporer, S.L. (1991). Deep-deeper-deepest. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 17, 323-333. doi:10.1037//0278-7393.17.2.323
- Sporer, S.L. (1993). Clothing as a contextual cue in face recognition. *German Journal of Psychology*, 17, 183-199. doi:10.1037/0278-7393.17.2.323
- Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 46A, 225-245. doi:10.1080/713756795
- Tanaka, J.W., & Sengco, J.A. (1997). Features and their configuration in face recognition. *Memory & Cognition*. 25, 583-592. doi:10.3758/BF03211301
- Thomas, S.M., & Jordan, T.R. (2002). Determining the influence of gaussian blurring on inversion effects with talking faces. *Perception & Psychophysics*, 64, 932-944. doi:10.3758/BF03196797



- Thompson, P. (1980) The Thatcher illusion. *Perception*, 9, 483-484.  
doi:10.1068/p090483
- Tredoux, C.G., Nunez, D.T., Oxtoby, O., & Prag, B. (2006). An evaluation of ID: an eigenface based construction system. *South African Computer Journal*, 37, 1-9.
- Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373. doi:10.1037/h0020071
- Woodhead, M.M., & Baddeley, A.D. (1981). Individual differences and memory for faces, pictures and words. *Memory and Cognition*, 9, 368-370.  
doi:10.3758/BF03197561
- Yasuda, T. (2005). Effects of configural information on face shape perception. *Japanese Journal of Psychology*, 76, 131-138.
- Young, A.W., Hay, D.C., McWeeny, K.H., Flude, B.M., & Ellis, A.W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14, 737-746. doi:10.1068/p140737
- Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configural information in face perception, *Perception*, 16, 747-759. doi:10.1037/0096-1523.26.2.527

## Figure and table captions

Figure 1. Blurring of external features illustrated on a synthesised face from EvoFIT. The level of Gaussian blur shown (from left to right) is zero (no blur), 4 cycles per f/w (low), 6 cycles per f/w (medium) and 8 cycles per f/w (high). In Experiment 1, participant-constructors constructed a single composite by selecting from arrays of 18 such faces with external features degraded at the same level. See Figure 2 for example composites constructed in this experiment.

Figure 2. Finished composites constructed in Experiment 1. These were produced by participant-constructors selecting from EvoFIT face arrays with one level of external-features blurring: (a) none, (b) low, (c) medium and (d) high. Image (e) is the target photograph shown to the constructors who subsequently constructed these composites (shown in colour in the actual experiment). Images (f) – (i) show internal features of (a) – (d), respectively. Participant-evaluators were asked to name the composites constructed in this experiment (some of which are shown here) either as complete images (top row) or as internal features (bottom row). After naming the composites, they named the target photographs (including the one shown above).

Figure 3. Images produced by individual constructors in Experiment 2 and presented to evaluators to name. They were produced using (a) no external features, or with external features containing (b) similar hair, (c) exact hair or (d) exact hair and shoulders. Image (e) is the target photograph (shown in colour in the experiment) used to construct these composites; its hair, and hair and shoulders, were extracted and presented to constructors for face synthesis in conditions (c) and (d), respectively. Unlike Experiment 1, no blurring was applied to the exterior facial parts during construction.

Figure 4. Example composites of UK footballer, Frank Lampard, produced from constructors in Experiment 3 in conditions of (left to right) normal-blur, extended-blur and internals-only. These and other images produced in each condition (along with four non-target composites) were shown to evaluators to name—the resulting mean naming scores (adjusted for target familiarity) can be seen in Table 3. Note that, unlike Experiment 2, constructors here were allowed to select hair; when doing so for this target, each person independently selected the same style (the constructor using extended-blur then lightened it).

Table 1. Percentage of complete and internal-features composites named correctly in Experiment 1, for composites constructed at different levels of external-features blur (none, low, medium or high). Values in parentheses are standard errors of the (by-item) means.

Table 2. Percentage correct naming of composites constructed in Experiment 2 using different types of external features: similar hair, exact hair, exact hair and shoulders, and no external features. For each of these conditions, the table lists naming of the complete face (top row) and naming of just the internal-features region (bottom row). Values in parentheses are standard errors of the (by-item) means.

Table 3. Composite quality following different presentation of external features at construction in Experiment 3. Values shown were calculated by dividing correct naming scores of composites by correct naming scores of targets, with the results expressed as percentages. Values in parentheses are standard errors of the (by-item) means.

## List of figures and tables

Figure 1.



Figure 2.

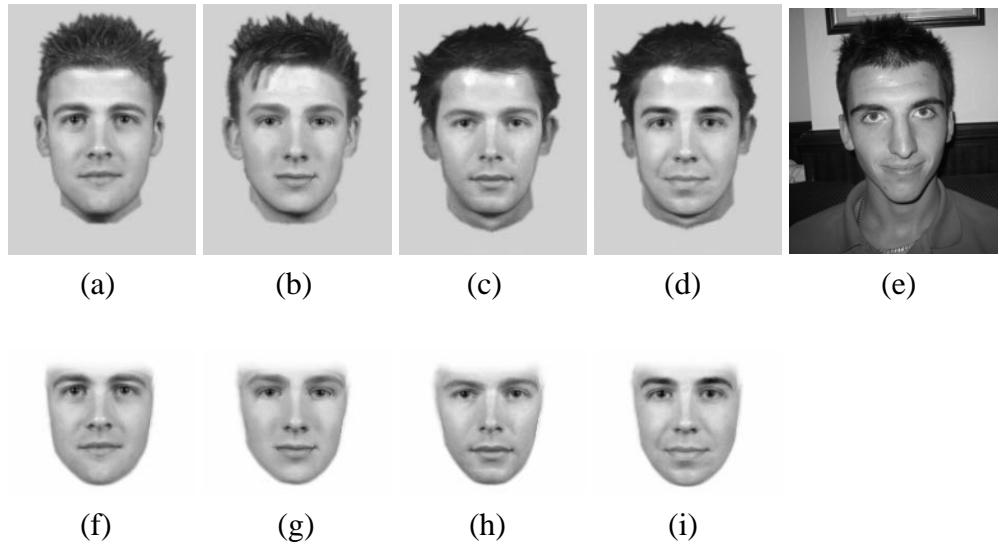


Figure 3.



Figure 4.



Table 1.

	Blur level				Mean
	None	Low	Medium	High	
Complete face	4.0 (1.8)	24.0 (4.5)	70.0 (5.5)	74.0 (2.1)	43.0 (2.3)
Internal features	4.0 (2.3)	10.7 (2.8)	54.0 (3.6)	69.3 (3.5)	34.5 (3.1)
Mean	4.0 (2.0)	17.3 (4.2)	62.0 (5.2)	71.7 (2.9)	38.8 (2.5)

*Note.* Numbers in parentheses are standard errors of the by-item means.



Table 2.

	Similar hair	Exact hair	Exact hair-and- shoulders	None (Internals-only)
Complete face	38.3 (4.2)	61.7 (3.3)	82.5 (3.2)	-
Internal features	4.2 (1.4)	13.3 (3.1)	39.2 (2.8)	48.3 (3.5)

*Note.* Numbers in parentheses are standard errors of the by-item means.

Table 3.

Target	Method of face construction		
	Normal blur	Extended blur	Internals only
Craig Bellamy	10.0	7.7	33.3
Michael Carrick	21.4	30.8	28.6
Joe Cole	25.0	37.5	43.8
Steven Gerrard	33.3	28.6	13.3
Frank Lampard	43.8	31.3	75.0
Gary Neville	15.4	80.0	50.0
John O'Shea	6.7	20.0	30.8
Paul Scholes	0.0	25.0	50.0
Alan Smith	40.0	81.3	81.3
John Terry	31.3	25.0	50.0
Mean	22.7 (4.6)	36.7* (7.7)	45.6* (6.6)

*Note.* Numbers in parentheses are standard errors of the by-item means.

\* Significantly different from Normal blur,  $p < .05$ .