Frowd CD, Jones S, Fodarella C, Skelton FC, Fields S, Williams A, Marsh J, Thorley R, Nelson L, Greenwood L, Date L, Kearley K, McIntyre AH & Hancock PJB (2014) Configural and featural information in facial-composite images, *Science and Justice, 54* (3). pp. 215-227.

This is the peer reviewed version of this article

Configural and featural information in facial-composite images

Charlie D. Frowd (1*)

Sharrome Jones (2)

Cristina Fodarella (3)

Faye Skelton (3)

Steve Fields (4)

Anna Williams (5)

John Marsh (3)

Rachel Thorley (3)

Laura Nelson (3)

Leah Greenwood (3)

Louisa Date (3)

Kevin Kearley (6)

Alex H. McIntyre (7)

Peter J.B. Hancock (7)

(1) University of Winchester

Winchester, SO22 4NR

* Corresponding author: Charlie Frowd, Department of Psychology, University of Winchester,
Winchester SO22 4NR, UK. Email: Charlie.Frowd@winchester.ac.uk. Phone: (01962) 624943.

(2) HMP Glenochil

Stirling, FK10 3AD

(3) School of Psychology

University of Central Lancashire, PR1 2HE

(4) HMP Edinburgh,

Edinburgh, EH11 3LN

(5) Cranfield University

Shrivenham, SN6 8LA

(6) Scientific Support

Lancashire Constabulary, PR4 5SB

(7) Department of Psychology

University of Stirling, FK9 4LA

Configural and featural information in facial-composite images

Eyewitnesses are often invited to construct a facial composite, an image created of the person they saw commit a crime that is used by law enforcement to locate criminal suspects. In the current paper, the effectiveness of composite images was investigated from traditional feature systems (E-FIT and PRO-fit), where participants (face constructors) selected individual features to build the face, and a more recent holistic system (EvoFIT), where they 'evolved' a composite by repeatedly selecting from arrays of complete faces. Further participants attempted to name these composites when seen as an unaltered image, or when blurred, rotated, linearly stretched or converted to a photographic negative. All of the manipulations tested reduced correct naming of the composites overall except (i) for a low level of blur, for which naming improved for holistic composites but reduced for feature composites, and (ii) for 100% linear stretch, for which a substantial naming advantage was observed. Results also indicated that both featural (facial elements) and configural (feature spacing) information was useful for recognition in both types of composite system, but highly-detailed information was more accurate in the feature-based than the holistic method. The naming advantage of linear stretch was replicated using a forensically more-practical procedure with observers viewing an unaltered composite sideways. The work is valuable to police practitioners and designers of facial-composite systems.
(218 words.)

Running head: Confeatural information in composites

Practitioner summary
The research investigates the effectiveness of facial-composite systems as used by police practitioners to locate people (offenders) who commit crime. It is found that different types of information were created differently for contrasting composite systems. A novel presentation format was also developed that should substantially improve recognition of real-world composites.
(50 words.)

1. Introduction

From a social perspective, faces are important visual stimuli: they are the main source of information regarding another person's identity, emotion and intention (e.g. Kemp, McManus & Pigott, 1990; Leder & Bruce, 2000; White, 2001). We are able to recognise familiar faces extremely well, despite transient changes in head pose (e.g. Bruce, Valentine & Baddeley, 1987; Moses, Ullman & Edelman, 1996) expression (e.g. Bruce, 1982; Patterson & Baddeley, 1977), lighting (e.g. Moses et al., 1996), and other such factors (for reviews, see Hancock, Bruce & Burton, 2000; and Johnston & Edmonds, 2009). Research (e.g. Ellis, Shepherd & Davies, 1979) has also indicated that recognition of familiar faces involves processing of *internal features* (the region comprising eyes, brows, nose and mouth) to a greater extent than *external features* (hair, face shape and ears). For a face to which we have had minimal exposure, a so-called *unfamiliar* face, transient changes in appearance exert a large influence on recognition (e.g. Bruce, 1982), as do external features (Ellis et al., 1979).

It is also the case that face recognition not only involves processing of facial features—eyes, brows, nose and mouth—but also the spatial arrangement or *configuration* of these features (for reviews, see Cabeza & Kato, 2000; and Farah, Wilson, Drain & Tanaka, 1998). In a prominent theoretical approach proposed by Diamond and Carey (1986), faces contain configuration first-order and second-order relational information. The former specifies typical spatial relations between features; for example, faces have two eyes above a nose, both of which are located above a mouth, etc. In contrast, second-order (sub-ordinate level) relational information describes the relative size of these spatial relations—simply, the distances between features for a specific face. In the current paper, we refer to this latter sense when describing configural processing of faces, and contrast it with recognition by processing of individual features (*featural* processing).

In an attempt to understand the contribution of featural and configural information for face recognition, of which both are important, and to provide insight into the processes which underlie recognition (Bruce, Burton & Craw, 1992), research has investigated the impact of changing the appearance of a face. Observers are typically presented with photographs of faces which contain image distortions such as photographic negation; Gaussian (blur) and other types of spatial filtering; caricature; rotation and inversion; pixelation; and various types of linear and non-linear stretch (e.g., Benson & Perrett, 1991; Hole, George & Dunsmore, 1999; Hole, George, Eaves & Rasek, 2002; Lander, Bruce & Hill, 2001). These kind of image distortions influence the appearance of the face as a whole and, with the possible exception of stretch, interfere with recognition; as such, they provide evidence that face recognition is a holistic process (hence the term *holistic face processing*). Other approaches similarly indicate the holistic nature of face recognition (e.g. Sinha & Poggio, 1996;

Young, Hellawell & Hay, 1987; Wells & Hryciw 1984). It is perhaps worth mentioning that featural and configural processing are not perceptually independent of each other (e.g. Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Yasuda, 2005). For example, if eye spacing is increased, a face appears to be wider, but if the mouth is enlarged, the distance between nose and mouth appears reduced.

In the following sections, we aim to provide an accessible review of four standard whole-face transformations that are frequently used in face-recognition research: photographic negation, picture-plane rotation, Gaussian filtering (blur) and linear stretch. We outline what these affine transformations reveal about processing of human faces and then, over the course of two experiments, investigate what they tell us about information contained in another class of stimuli, facial composites. Composites are normally constructed from eyewitness memory using one of a number of standard composite systems. An understanding of what information is rendered accurately in composites, in particular featural and configural information, is not only theoretically interesting but also offers the possibility of improving the effectiveness of composite images. See Figure 1 for example transformations and Table 1 for a summary of feature and holistic information which is likely to be used by observers when recognising images seen with these transformations.

Table 1 about here

1.1. Photographic negation.

Photographic negation involves reversing the contrast polarity of an image. The transformation was assessed systematically by Galper (1970) and has since provoked extensive research (e.g., see Russell et al., 2006). The research has produced consistent findings, and the general consensus is that faces seen in photographic negative are difficult but not impossible to recognise (e.g. Bruce & Langton, 1994; Hole et al., 1999; Kemp et al., 1990; Liu, Collin, Burton & Chaudhuri, 1999; Phillips, 1972). The work thus indicates that some information which is useful for recognition is disrupted with negation.

One explanation of the photographic-negative effect relates to disruption in the processing of pigmentation (e.g., Bruce & Langton, 1994; Hole et al., 1999; Russell et al., 2006; Vuong et al., 2005), such as when the colour of hair and eyes are altered. Negation may also interfere with the pattern of shading across a face (e.g. Johnston, Hill & Carman, 1992). There is an interaction between the shape of the face and direction of lighting (Russell et al., 2006; Vuong et al., 2005),

known as shape-from-shading relations, which can be used to estimate the three-dimensional structure of a face (Phillips, 1971); irregular patterns of shading (with negation) can impair formation of this mental representation (Hill & Bruce, 1996). It is even possible that negation may interfere with global processing of the face. Kemp et al. (1990) found that people's ability to detect a small horizontal or vertical displacement of the eyes was disrupted with negation, implying that configural (second-order relational) processing was being impaired under this transformation.

1.2. Rotation and inversion.

Faces that are inverted—that is, turned upside down—are disproportionately difficult to recognise compared with inversion of other objects (Yin, 1969; for reviews, see Valentine & Bruce, 1988, and Tanaka & Gordon, 2011), an effect that is attributed to the disruption of configural relations (e.g. Carbon & Leder, 2005; Freire, Lee & Symons, 2000; Leder & Bruce, 1998). For example, Searcy and Bartlett (1996) manipulated faces by changing local features (such as blackening teeth) or by distorting the facial configuration (such as moving eyes apart). Local feature manipulations made faces appear more grotesque in both upright and inverted orientations, but configural manipulations affected the perception of upright but not inverted faces. The work underscores the importance of configural relations for processing faces in their normal, upright orientation.

In more detail, Goffaux and Rossion (2007) found that inversion affects vertical-relational information (e.g. vertical position of mouth) to a greater extent than either horizontal-relational information (e.g. eye spacing) or a change of feature information (e.g. blackening teeth). Indeed, Dakin and Watt (2009) suggest that facial identity is conveyed by a top-to-bottom oriented structure, a one dimensional vertical 'bar code' that is disrupted with inversion. Rossion (2008) also argues that local relational information is maintained with inversion, in particular in the horizontal direction around the eyes, while vertical relations tend to be disrupted, particularly around the mouth and for relations which involve longer distances (e.g. from brows to nose). Tanaka and Gordon (2011) also find evidence of greater disruption with inversion for lower relative to upper half of a face, especially when making featural and relational discriminations around the mouth region.

Research has also considered rotation in the picture plane at angles other than 180 degrees. Results reveal disruption to recognition, although usually to a lesser extent than for inversion. Valentine and Bruce (1988) found a linear trend between degree of disorientation and response time (see also Bruyer, Galvez & Prairial, 1993; Schwaninger & Mast, 2005). Collishaw and Hole (2002) observed that configural processing becomes more disrupted the further faces are oriented

away from upright. Other research using different paradigms has reported departures from linearity at around 90° (e.g. Jacques & Rossion, 2007; Rossion & Boremanse, 2008; but see also Lewis, 2001).

1.3. Stretch.

Another useful transformation for understanding the importance of information in an image is stretch. When either the height or width is increased linearly, configural properties are disrupted: absolute distances are changed, such as the length of the nose, as well as relative distances, such as the ratio between eye separation and nose length. This manipulation obviously also changes the physical size of individual features in either the relevant horizontal or vertical direction.

Hole et al. (2002) reported that recognition accuracy and reaction time for familiar faces were unaffected overall by linear stretch in the vertical or horizontal plane, even for manipulations of up to 200%. The work indicates the robust nature of this (null) effect, although, in one of their studies (Experiment 1), there was a small but reliable decrement in accuracy for horizontal stretch. Subsequent research has also found equivalent effects between stretched and non-stretched faces (Bindemann et al., 2008; Hole, 2011).

This result is extraordinary given the large physical change which occurs when an image is stretched. It also raises the question as to how linearly-stretched faces are recognised (see following section); and, why similar effects are observed in both the horizontal and vertical plane (although the former may be slightly more disruptive to recognition than the latter). Hole et al. (2002) have also compared recognition accuracy for globally and non-globally stretched faces, the latter only stretching one half of the face. Non-global vertical stretch impaired recognition, which suggests that configural cues were being interfered with from this altered portion of the face.

1.4. Gaussian blur.

Our cognitive system processes facial images at different spatial scales. This phenomenon has been investigated with images filtered using pixelisation, and low-, high- and band-pass filtering (e.g. Costen, Parker & Craw, 1996; Sergent, 1984). While high-pass filtering tends to retain both featural and configural information, low-pass or Gaussian (blur) filtering conceals detail of hair, eyes, brows, etc. For example, Lander et al. (2001) observed that recognition accuracy decreased with increasing blur (by condition) for famous-face images. Sergent (1984) argues that blurring degrades individual

features more rapidly and fully than configural relations: obviously, it is difficult to perceive facial features in detail when blurred, although their placement on the face remains intact.

Gaussian blur also conceals facial features when these parts are presented in an unusual format. Schwaninger, Lobmaier and Collishaw (2002) presented intact faces, or faces where the position of features had been scrambled—to create 'invalid' first-order configurations. Faces that had been either blurred or scrambled reduced recognition relative to unaltered images, although this level remained above chance. However, blurring reduced recognition of scrambled faces to chance, suggesting that this transformation removes [featural] cues emerging from facial features.

In further work, Hole et al. (2002) found that blurred faces remain recognisable even when linearly stretched. Although recognition was poor under both stretch and blur, faces remained recognisable to some extent. They interpreted these data as evidence that recognition of stretched faces is based on configural cues rather than on local (feature) detail.

1.5. Facial composites.

In a forensic setting, a facial composite is a visual likeness of an offender (for a review, see Frowd, Bruce & Hancock, 2009). It is usual for witnesses (who could also be victims) of serious crime to describe and construct a composite of the offender's face they have seen, typically an unfamiliar identity. This procedure is carried out with the help of a forensic artist using pencils or crayons, or a police practitioner using specialized computer software. The resulting composite images are circulated within a police force or more generally as part of a public appeal for information, often appearing in newspapers, on wanted-person's webpages and on TV crime programmes. The aim is that someone (a police officer or member of the public) who is familiar with the offender will name the composite and so generate a suspect on whom police can focus their enquiries.

There are two contrasting software systems for witnesses to construct a composite (Frowd et al., 2005b, 2012c). First are traditional 'feature' systems such as PRO-fit and E-FIT in the UK, and FACES and Identikit 2000 in the USA. Witnesses describe the appearance of an offender's face to a forensic practitioner, who then presents facial features (hair, eyes, nose, etc.) to match. Since facial features are identified better when seen in a whole-face than an isolated context (e.g. Davies & Christie, 1982; Tanaka & Farah, 1993), feature choices are presented in an intact face; witnesses also alter size and position of features with the aim of producing the best likeness. They may also work with a sketch artist, a practitioner who draws the face by hand using a feature-by-feature procedure.

In either case, the task is principally based on local feature elements and so one would anticipate that the resulting composites would be more accurate in terms of featural than configural cues.

A second type of system involves selection of intact (whole) faces. To our knowledge, there are three such 'holistic' systems which have been developed extensively: EFIT-V (Gibson, Solomon & Pallares-Bejarano, 2005) and EvoFIT (Frowd, Hancock & Carson, 2004) in the UK, and ID (Tredoux, Nunez, Oxtoby & Prag, 2006) in South Africa. Witnesses describe the offender's face and repeatedly select from arrays of faces, with the software 'breeding' choices together, to allow a composite to be 'evolved'. Observers are asked to base selections on the overall appearance of the face, not individual features, with the aim of producing faces that are accurate in terms of global properties, especially configural cues. Given the importance of holistic information for recognition, such composites should be recognisable by other people. For the same reason, holistic systems allow users to alter global properties of an evolved face (for EvoFIT: age, weight, masculinity, health, etc.).

In terms of effectiveness, the holistic EFIT-V and the feature systems appear to perform similarly (Brace, Pike & Kemp, 2000; Bruce et al., 2002; Davies, van der Willik & Morrison, 2000; Frowd et al., 2005b, 2011; Valentine et al., 2010), with composites constructed from a recent memory achieving approximately 25% correct naming by other people. When tested under similar conditions, EvoFIT composites are named with higher accuracy; in a recent version of the system, similar to that used here, performance was measured at 54% correct (Frowd et al., 2011).

1.6. Current experiments.

Two experiments are presented which investigate our ability to correctly name composites seen in standard formats in face-recognition research. In Experiment 1, participants viewed composites of familiar faces as veridical (unaltered) images, or when these images were rotated, linearly stretched, blurred or photographically negated. Composites were used from two feature systems, PRO-fit and E-FIT, and from the holistic EvoFIT. These manipulations were expected to reduce people's ability to correctly name the face relative to veridical composites, except for stretch, for which no change was expected. Also, as indicated by high blur and inversion, featural information was likely to be more accurate in composites from feature than holistic systems, with the opposite effect by system for configural information. In Experiment 2, we attempt a replication of the surprising result from Experiment 1 that stretch substantially improved composite naming—a result that would appear to be valuable to forensic practitioners. We note that the stretch technique has now been successfully replicated for a second time (Frowd et al., 2013), as described in Section 4.4.

2. Experiment 1: Standard Image manipulations applied to facial composites

2.1. Method

2.1.1. Design. The first experiment manipulated five main factors in a mixed-factorial design.

 Factors. The first factor was presentation type. Participants saw a series of composite faces to name in one of the following preparations: veridical, negation, blur, rotation and stretch. See Materials (below) which describe how the composites were constructed and prepared for presentation to participants. While the veridical condition contained unaltered composites, as created by the face constructors, and negation was contrast polarity reversal, the others (blur, rotation and stretch) contained conditions with two or more levels.

To investigate the accuracy of configural information, composites were given one of three levels of Gaussian blur, similar to Lander et al. (2001). The level of image distortion was minor, moderate and strong; these levels respectively correspond to what we refer to as *low*, *medium* and *high* blur (see Materials for image settings). The expectation was that progressively fewer feature composites would be correctly named from conditions of low to medium to high; however, there should be little difference by condition for holistic composites. Conversely, for rotation, composites were presented at 90° from upright in one condition and 180° in the other. Perceiving these composites with increasing rotation (from 0° to 90° to 180°) was expected to result in observers processing the faces to a greater extent by individual features, and have the opposite effect to blur: increasing rotation should increasingly interfere with recognition of holistic composites.

For participants assigned to the 90° rotation condition, half of them saw composites rotated to the right and the other half to the left. This was done more for experimental completeness, to produce balanced stimuli sets, and we did not expect there to be a bias for composites seen rotated in one direction or the other. Similarly, the width of composites was non-proportionally increased by 100% for a stretch-horizontal set and decreased by 100% for a stretch-vertical set; participants saw one of these two sets. As for 90° rotations, these two linearly-stretched conditions were not expected to differ reliably from veridical or from each other—although, if a difference were to be observed, it was predicted that horizontal would show a decrement in naming (as suggested by the mild recognition deficit for horizontal relative to veridical in Experiment 1, Hole et al., 2002).

To summarise, there were eight main presentation conditions: veridical, photographic negative, Gaussian blur (low, medium and high), rotation (90° and 180°) and stretch. Two of these conditions were subdivided: 90° rotation (right and left) and stretch (vertical and horizontal).

The second experimental factor was the type of composite system used for constructing the face. The systems used were based on availability, and were popular feature and holistic systems. For the feature type, composites were available from E-FIT and PRO-fit; as these systems produce equally-identifiable composites (e.g. Frowd et al., 2005a, 2005b), images were chosen arbitrarily from one system or the other. For the holistic system, composites were used from EvoFIT. All composites were of different identities. In the study, participants inspected composites from one system and then the other in a blocked design that was counterbalanced across participants. System (feature and holistic) was therefore a within-subjects' factor.

This blocked design also permitted an investigation into whether viewing composites from one type of system would interfere with recognition of composites from the other. To our knowledge, this possibility has not been formally assessed, but, in the real world, observers may attempt to recognise composites from different production methods and so an understanding of this issue is worthwhile. As naming is a process which operates on the face as a whole, it is conceivable that composites which are accurate in one respect may result in subsequent face processing of that type which influences recognition (either positively or negatively). Indeed, there is evidence to suggest that face processing can be biased, with a knock-on effect on recognition (e.g. Lewis, 2006; Schooler, 2002). In our case, looking at composites from a feature system (in the first block), believed to be more accurate in terms of individual features, was expected to promote a processing bias towards individual features and reduce naming when holistic composites were seen (in the second block): in contrast, when holistic composites were seen first, this would result in a holistic bias and facilitate naming of feature composites (seen second). The third experimental factor was block order, between-subjects.

To be able to appropriately interpret results for composites, it is important to verify how the manipulations influenced recognition of stimuli that were not prone to error (like composites). So, after the composites had been named in the assigned presentation, the photographs originally used to construct the composites were presented in the same way. For example, if participants saw composites rotated at 180°, they would then be presented with photographs also rotated at 180°. These photographs were shown in greyscale since, with colour, participants might have used pigmentation cues for recognition (e.g. Yip & Sinha, 2002). This decision also fits nicely with the customary image mode for composites, greyscale, as is the case here.

When the study was designed, however, we were concerned that people may struggle to correctly name the composites, especially in the more challenging high blur and 180° rotation conditions, which would potentially lead to low experimental power and inconclusive results. Three

steps were taken to address this possibility. Firstly, composites were selected with a known level of accuracy. As outlined in the Materials, each composite was extracted from a past research project if it had been correctly identified in a spontaneous-naming task; therefore, naming should be possible here. Secondly, after spontaneous composite naming, composites were shown a further three times (in the assigned presentation) along with a semantic cue designed to increasingly facilitate recognition. As it turned out, spontaneous-naming data were sufficient and thus these 'cued' data were not formally analysed. It is perhaps worth emphasising that spontaneous naming is the most forensically-relevant measure, and it is on these data that our results are based. Thirdly, participants would not be able to correctly name the composites if they were unfamiliar with the relevant identities. So, as a final exercise after naming of composites (spontaneous and cued) and target pictures (also in the assigned presentation), participants were asked to name the unaltered target pictures (images used for face construction). An *a priori* rule was applied: participants must correctly name at least 75% of the target pictures for their data to be included in the analysis. For each participant who did not meet this criterion, another person was recruited in that condition. As a consequence of this design, in the veridical-presentation condition, veridical photographs were presented twice: once after composite naming and once again for the target-familiarity check.

Summary. The design was 8 (between-subjects) presentation [veridical, photographic negative, blur (low, medium and high), rotation (90 and 180 degrees), and stretch] x 2 (within-subjects) composite system (feature and holistic) x 2 (between-subjects) block order of composites (first and second) x 2 (within-subjects) image type (composite and photograph). Within presentation, there were two levels for 90° rotation (left and right) and two levels for stretch (vertical and horizontal). The design also included semantic cue as a factor, composite-naming data that were not analysed. Participants first named composites (blocked by system) for one type of presentation, followed by target pictures presented in the same way as composites, and finally veridical target pictures.

Predictions. Based on the above research for photographs of faces, presentations were expected to reduce correct naming of composites relative to veridical except for stretch, for which no change was expected, and for low blur, as this change was very subtle. Manipulations were thought to produce the same pattern of effects for naming of target photographs. For composites, some conditions were likely to affect naming of one type of system more than others, and so the interest was also in differences between systems by presentation. In the following three main predictions, we assume that holistic composites have better configural than featural information, but feature composites have better featural than configural information.

Firstly, increasing the angle of rotation for a face principally results in reduced processing of configural cues and greater processing of individual features. So, observers who name a composite image that is rotated have a tendency to use featural information. As rotation increases, here from 0° (veridical) to 90° to 180°, naming of holistic composites should be increasingly affected. In fact, as 90° is midway between the other two levels, and produces intermediate performance (e.g. Valentine & Bruce, 1988), we predicted a linear decline in naming across these categories. In contrast, feature composites should not be affected. Next, while there are various explanations of the recognition deficit for photographic negatives, all appear to be related to disruption to the overall appearance of the face—interference to pigmentation and/or configural cues. As such, recognition of negatives is likely to be driven by individual features, and so the prediction for negation is the same as for rotation: holistic composites should be affected to a greater extent than feature composites.

Secondly, higher levels of blur applied to a face increasingly conceal individual features and processing increasingly involves configural relations. As medium blur was designed as a stimuli to be approximately midway between low and high, increasing blur from low to medium to high should result in a decreasing linear trend for naming of feature composites—that is, information useful for recognition of these composites is progressively concealed. In contrast, holistic composites should be unaffected. For linear stretch, face processing is also thought to be largely configural (e.g. Hole et al., 2002) and so the same outcome was expected for naming of stretched stimuli.

Finally, as described in the Materials below, composites were selected to give mean naming that was approximately the same between systems (feature or holistic), and so an overall difference by system was not expected for veridical presentation. However, as composites from one type of system should influence recognition of composites from the other, block order was predicted to interact with system: feature composites would demote naming of holistic composites, while holistic composites would facilitate naming of feature composites.

2.1.2. Materials. Face-recognition research has often involved photographs of celebrities as stimuli (e.g. Burton, Jenkins, Hancock & White, 2005; Collishaw & Hole, 2002; Davies et al., 1978; Hole et al., 2002; Lander et al., 2001; Valentine & Bruce, 1988). In the present research, we followed this approach and involved composites constructed of celebrity faces. While this has the obvious advantage of facilitating both participant recruitment and generalisation of results with the earlier studies, in criminal investigations, faces are obviously not normally of celebrities. There is evidence, however, that the type of face, celebrity or non-celebrity, does not influence composite naming. Consider a design involving participants looking at a photograph of a target face and, either

immediately or a few hours later, describing and constructing a composite of it. Using this design, Frowd et al. (2004) found that people named celebrity E-FIT composites with a mean of 17% correct; it was 18% in Frowd et al. (2005b) (combined mean for E-FIT and PRO-fit composites) and 25% in Brace et al. (2000) (E-FIT composites). For PRO-fit composites of personally-familiar (non-celebrity) faces, Bruce et al. (2002) measured correct naming of 19% (Stage 3: average of best and worst individual composite) and Frowd, Bruce, McIntyre and Hancock (2007a) at a similar level, 22%. Naming therefore appears to be fairly consistent for celebrity and non-celebrity composites.

In spite of this reasonably-consistent performance by target category, there are large individual differences in people's ability to construct a recognisable image. In fact, some composites are of such poor quality that they are rarely if ever correctly named (e.g. Frowd et al., 2005a, 2007c), a situation, as mentioned above, which would lead to low experimental power and ineffective analyses for the current study. We overcame this issue by selecting celebrity images which people had correctly named from previous projects (Frowd et al., 2004, 2005b, 2007b). Using this approach, all of the holistic composites available were of a target which were familiar to participants: for the feature composites, we were able to select half that were familiar and half that were unfamiliar to the person constructing the face—a situation that is more common forensically. This design allowed familiarity to be factored in a regression analysis to see if support could be found for existing research which indicates that composites are more identifiable when constructed of a familiar identity (Davies et al., 2000; Frowd et al., 2011; Koehn and Fisher 1997; Wogalter & Marwitz, 1991).

There are also large individual differences in peoples' ability to correctly name composites (e.g. Frowd et al., 2007c), and so it is sensible to consider experimental power and effect size. The aim was to detect a medium effect size, to be of practical significance, and this was achieved by estimating the number of composites required per condition. The estimation was based on a planned Mixed-Factorial Analysis of Variance for the two main factors, system and presentation. A G*Power calculation (Faul, Erdfelder, Lang & Buchner, 2007) revealed that 10 targets (items) per condition would be sufficient to detect a within-between subjects interaction; parameter settings: effect size f = 0.25, α = .05, Power 1-β = .8, N (groups) = 2, N (measurements) = 8, r (repeated measures) = .7 and non-sphericity correction ε = 1.

Ten celebrity composites were thus selected from archives for both the holistic system (EvoFIT) and the feature system (E-FIT and PRO-fit). Images were chosen such that mean naming was fairly-good and approximately equal for each type of system: 25% correct. Identities were different throughout, allowing all 20 items to be presented to each participant, blocked by system.

Pilot testing revealed that the procedure took around 20 minutes to complete per person, a typical interval for psychological testing, meaning that fatigue-related effects were unlikely.

Composites and (greyscale) target photographs were cropped to dimensions of 360 x 480 pixels, and eight main manipulations were carried out on each image. Adobe Photoshop's contrast-inversion filter was used to prepare photographic negatives, and the Gaussian filter for 'low' (2 cycles per face width), 'medium' (4 cycles/f-w) and 'high' (8 cycles/f-w) blur. Otherwise, Microsoft Word was used to present images at 90° and 180° rotation, and also to resize them for the 100% stretch conditions. Stimuli were printed to veridical-image dimensions of 6 cm (wide) x 8 cm (high). See Figure 1 for example manipulations.

Figure 1 about here.

2.1.3. Participants. Participants were an opportunity sample of 96 volunteer staff and students from the University of Central Lancashire, Preston. There were 51 males and 45 females with an age range from 18 to 53 (M = 24.9, SD = 7.4) years. Twelve people were assigned to each of the eight main levels of the between-subjects' factor, presentation (veridical, negation, low blur, medium blur, high blur, 90° rotation, 180° rotation and stretch); within this allocation, for 90° rotation, six participants were each assigned to 90° right and 90° left, and for stretch, six were each assigned to horizontal and vertical. Also, for the between-subjects' factor of block order, 48 participants saw holistic composites first while the other 48 saw feature composites first. Each person was recruited on the basis of being generally familiar with celebrities. Two participants did not meet the *a priori* rule relating to target familiarity, and were replaced, to give the sample described here.

2.1.4. Procedure. Participants were tested individually and tasks were self-paced. Each person was told that they would be shown facial composites of well-known celebrities to name, or to guess if unsure: participants could also choose not to offer a name. They were randomly assigned, with equal sampling, to see composites from one type of presentation (veridical, negation, 90° rotation, 180° rotation, low blur, medium blur, high blur and stretch). For rotate 90° and stretch conditions, they saw either rotate left or rotate right for the former, with equal assignment, and either vertical or horizontal stretch for the latter, equally assigned. Composite system was blocked such that the 10 holistic composites were presented first for half of the participants (followed by the 10 feature

composites), with random assignment, and the 10 feature composites first for the other half (followed by the 10 from the holistic system). Composites from the assigned presentation were shown sequentially, from the assigned system and then from the other, and participants said a name for the face (or not) as instructed. Images were presented at a fixed viewing distance (approx. 40cm); for participants inspecting rotated composites, they were requested not to turn their head to the side to view the face (which may have otherwise reduced the effect of rotation). Images were shown in a different random order of presentation for each person. No feedback was given as to the accuracy of responses, the same as in the following tasks [1].

Next, participants assigned to conditions except for veridical were told that they would name the photographs used to construct the composites, and that these images would be seen in the same way (i.e. with negation, rotation, blur or stretch). Target pictures were blocked by system in the same order as for composites. Finally, as part of a target-familiarity check, participants were asked to name the target photographs as initially seen by the face constructors. The procedure for participants assigned to veridical was slightly different (as target and composite images were already veridical): these participants were first asked to name the relevant target photos, and then (for the target-familiarity check) to have another go at naming them. In all cases, images were presented sequentially and in a different random order for each person.

Participants were debriefed as to the purpose of the experiment.

2.2. Results

Four sets of responses were collected from participants: (1) spontaneous naming of composites (at eight main levels of presentation), (2) three sets of responses for composites at each level of semantic cue (not formally analysed), (3) spontaneous naming of targets (at eight main levels of presentation) and (4) spontaneous naming of veridical targets. Participant responses were checked for missing data (of which no cases were found) and scored for accuracy with respect to the relevant identity: a composite or target photograph was given a numeric value of 1 if participants gave either the correct name of the person or a description that unambiguously identified him (an approach used in past research: Bruce, Hanna, Dench, Healey & Burton, 1992), or a value of 0 otherwise.

---

[1] In the next part, participants were invited to name the composites again, although this time information would be provided about the relevant identity. Images were presented again, in the same order as before, and the researcher read aloud the semantic cue for the composite. This cued-naming procedure was repeated twice more, each time with a more-specific semantic cue. (It is these data that we did not need to analyse.)

Two main analyses are presented below. In the first, correct naming scores of composites were subjected to Analysis of Variance (ANOVA) to explore group differences. As it is important to know how the manipulations affect faces in general, the naming data for composites were compared against control data: naming scores collected for manipulations on target photographs. In the second analysis, linear regression was used to further indicate which information, featural or configural, was responsible for naming of veridical, negative and stretched composites.

2.2.1. Naming of composites and target photographs. Correct naming of veridical photographs in the target-familiarity check was very-high overall (M = 88.3%, SD = 3.1%), indicating that participants appropriately had very-good familiarity with the relevant identities. The range of these mean scores by condition was somewhat variable (81.7% < M < 94.2%) and, to avoid chance differences in target familiarity, we computed 'conditional' naming scores for spontaneous naming of composites—a standard measure used in the discipline (e.g. Frowd et al., 2005b). Conditional naming scores were computed as the number of correct responses given for each composite divided by the number of correct responses given for the relevant veridical photograph shown during the target-familiarity check. As we were also interested in how the manipulations affected target photographs, the same measure was used for targets: the number correct for target photographs (by presentation) divided by the number correct for the relevant veridical photograph (from the target-familiarity check).

Computation of conditional naming scores in this way naturally leads to a by-items type of analysis, which is our preference for studying group effects over by-participants analysis (as the latter tends to suffer from outlier effects). Mean conditional responses by image and presentation types are shown in Table 2. For this initial summary, data are shown averaged by system (holistic and feature), block order of presentation (first and second), and sub-condition for 90° rotation (left and right) and for stretch (vertical and horizontal).

Table 2 about here

Mixed-Factorial ANOVA was conducted on conditional naming scores for presentation (veridical, negation, 90°, 180°, low blur, medium blur, high blur and stretch), block order (first and second), image (composites and photographs) and system (holistic and feature). Note for this

by-items analysis, identities which are repeated across conditions are treated as within-subjects factors; here, within-subjects factors are presentation, order and image: system is between-subjects.

Mauchly's Test of Sphericity was significant for presentation [W = 0.07, X^2(27) = 41.5, p = .042, ε = .59] and so degrees of freedom for this factor (below) were corrected using Greenhouse-Geisser. ANOVA was significant for both image [F(1,18) = 517.6, p < .001, η_p^2 = .97], as naming of target pictures (M = 90.5%, SD = 5.8%) was appropriately very-much higher overall than naming of composites (M =24.9%, SD = 14.5%), and presentation [F(4.1,74.3) = 8.8, p < .001, η_p^2 = .33]. Relative to veridical, two-tailed simple contrasts of the ANOVA indicated a naming decrement (p < .05) for negation, medium blur, high blur, and image rotations at 90 and 180 degrees, but a surprising naming advantage for stretch (p = .033). System was non-significant [F(1,18) = 0.6, p = .47].

The analysis revealed three reliable interactions involving system: order x system, order x presentation x system and the four-way interaction (Fs > 2.4). We now follow-up on these significant results (below) by conducting separate analyses, first to explore consistency of naming for veridical composites and target images, and then by considering each type of manipulation in turn: negation, rotation, blur and stretch. For brevity, statistics are not included for the main effect of image, which was always highly significant (p < .001), nor for the main effects of system and order, which were always non-significant (p > .05): only results from presentation and its interactions are reported.

2.2.2. Veridical images. Conditional naming scores for veridical composites and veridical targets were subjected to ANOVA. As mentioned above, the main effect of image type was significant (p < .001), and the two main effects (order and system) and their various interactions (including image type) were non-significant (ps > .05). This result itself indicates that, as designed, holistic and feature composites were not significantly different to each other. A lack of reliable interactions involving image type indicates that naming of target photographs, as for composites, was consistent by block order and system. This result also indicates that, contrary to expectation, naming of veridical composites from one system (feature or holistic) does not interfere with naming of veridical composites from the other (holistic or feature)—although, as reported below, such a bias is evident for other types of presentation.

2.2.3. Negation. Conditional naming scores were next analysed by 2 (presentation: veridical and negation) x 2 (image: composites and targets) x 2 (block order: first and second) x 2 (system: holistic and feature) ANOVA. Presentation was significant [F(1,18) = 7.5, p = .014, η_p^2 = .29], indicating that

negation led to a reliable decrement in overall naming for both composites and target photographs. Order x image was also significant [$F(1,18) = 5.2$, $p = .034$, $\eta_p^2 = .23$], and a Simple-Main Effects analysis revealed that the interaction emerged as naming increased from the first to the second block for composites ($p = .005$, Cohen's $d = 0.70$) but remained the same for target photographs ($p = .48$). So, as expected, negation interfered with naming of composites and target pictures; for composites, people improved in the second block, arguably because viewing composites as photographic negatives is a rather unusual combination that takes some getting used to. Also, negation did not reveal processing differences between holistic and feature composites.

2.2.4. Rotation. Mean conditional naming scores were somewhat lower from veridical to 90°, and also from 90° to 180°. These data were analysed by 3 (presentation: veridical, 90° and 180°) x 2 (image: composite and target photograph) x 2 (order: first and second) x 2 (system: holistic and feature) ANOVA. Rotation was significant [$F(2,36) = 7.0$, $p = .003$, $\eta_p^2 = .28$] and a planned polynomial contrast (in the order of veridical, 90° and 180°) revealed a significant linear ($p = .003$, $\eta_p^2 = .39$) but not quadratic trend ($p = .81$). Presentation x system was not significant which indicates that increasing rotation led to progressively worse naming for both systems, not just for EvoFIT; it also indicates that the gradient (slope) was the same for image type, implying that rotation negatively affects these two types of stimuli (composites and target photographs) equivalently.

Block order x system was reliable [$F(1,18) = 10.6$, $p = .004$, $\eta_p^2 = .37$]. As illustrated in Table 3, for the first block, holistic composites were named somewhat less often than feature composites ($MD = 9.3\%$), but Simple Effects indicated that this comparison was not significant ($p = .11$). However, holistic images were named reliably more often viewed in the second than the first block ($p = .030$, $d = 0.82$); also, while feature composites indicated the reverse effect, this contrast fell short of significance ($p = .08$). So, for rotation, a manipulation which tends to involve processing of individual features, this result suggests that processing of composites believed to be more accurate by features (E-FITs and PRO-fits) carries over to processing of composites believed to be more accurate in terms of configural relations (EvoFITs). (The associated finding of an approaching significant effect in the opposite direction for feature composites argues against this result simply being caused by practice.)

Table 3 about here

We next explored if images rotated at 90° to the left or the right were named differently—for which we predicted a null result. The analysis was 2 (orientation: 90° left and 90° right) x 2 (order: first and second) x 2 (system: holistic and feature) x 2 (image: composite and photograph) ANOVA. Orientation was unexpectedly significant [$F(1,18) = 8.5$, $p = .009$, $\eta_p^2 = .32$], as images presented rotated to the right ($M = 61.0\%$) enjoyed higher naming than images rotated to the left ($M = 53.5\%$). This curious result could be caused by differences in the hemispheric processing of faces (e.g. Schooler, 2002) and would appear to be worthy of further investigation. Block order x system was reliable [$F(1,18) = 6.8$, $p = .018$, $\eta_p^2 = .27$], as found above.

So, increasing rotation interfered equally with naming of target pictures and composites from both types of system. This indicates that holistic composites and, contrary to expectation, feature composites are harder to recognise with rotation. Also, as rotated composites—especially at the most extreme manipulation of 180°—are still named fairly-well for the two systems, featural information is valuable to some extent in both types of composite. In addition, there was some evidence that rotation differentially interfered with naming of different types of composites.

2.2.5. Blur. Gaussian (blur) filtering conceals information about individual facial features. It was expected that increasing levels of blur would increasingly reduce naming of both feature composites and target photographs; holistic composites would be unaffected. The initial analysis did not include veridical composites and was 3 (presentation: low, medium and high blur) x 2 (image: composite and target photograph) x 2 (order: first and second) x 2 (system: holistic and feature) ANOVA. Reliable effects emerged for image x system ($p = .02$), image x presentation ($p = .018$), order x presentation x system ($p < .001$) and the four-way interaction ($p = .016$). Based on predictions, analyses were followed-up separately for target photographs and both types of composites, as follows.

For target photos, 3 (presentation: low, medium and high blur) x 2 (order: first and second) x 2 (system: holistic and feature) ANOVA was significant only for presentation [$F(2, 36) = 11.3$, $p < .001$, $\eta_p^2 = .39$]. Polynomial contrasts revealed a significant linear ($p = .001$, $\eta_p^2 = .46$) but not quadratic ($p = .67$) trend in the predicted order: higher levels of blur gave rise to lower levels of correct naming. There was also no significant difference between veridical and low-blur target photographs [$t(19) = 0.4$, $p = .69$]. For composites, ANOVA of these same factors was significant for both presentation [$F(2, 36) = 11.3$, $p < .001$, $\eta_p^2 = .39$] and the three-way interaction [$F(2,36) = 10.7$, $p < .001$, $\eta_p^2 = .37$]. In the remainder of this section, we analyse this interaction for composites presented in the first and then the second block. See Table 4 for a summary.

For composites presented in the first block, presentation x system was reliable [$F(2,36) = 7.8$, $p = .002$, $\eta_p^2 = .30$], and so this interaction was explored by polynomial trends—as feature composites were expected to show a linear decrease with increasing (categorical) blur, with no change for holistic composites. For feature composites, there was indeed a reliable linear trend ($p = .047$, $\eta_p^2 = .37$), but this was in the opposite direction, an *increase* in naming with increasing blur; it was also weak as the quadratic component was sizeable and approached significance ($p = .084$, $\eta_p^2 = .30$). The result is understandable, as the means for low and medium blur are virtually the same (i.e. for a linear trend, medium should be about midway between the two other categories). Polynomial contrasts are thus not the best way to analyse these data. A more-acceptable analysis, however, involves simple contrasts of the ANOVA: low ($p = .013$, $d = 0.89$) and medium ($p = .047$, $d = 1.3$) blur were inferior to high; there was no significant difference between low and medium [$t(9) = 0.8$, $p = .24$]. For holistic composites, there was another unexpected result: the pattern of means suggest a sizeable linear decline. The trend is indeed linear ($p = .015$, $\eta_p^2 = .51$), but not quadratic ($p = .90$). So, the negative influence of increasing blur on naming is linear for both holistic composites and target photographs. In a follow-up analysis, blur (low, medium and high) and image (holistic composites and target photographs) do not interact with each other [$F(2,18) = 0.3$, $p = .80$], and thus the gradients of these two decreasing lines do not differ reliably. So, while correct naming levels are considerably higher for target photographs than holistic composites, the effect of increasing blur is equivalent between these two types of image.

Table 4 about here

We next consider the effect of blur relative to veridical composites. The table hints at an interaction, with low blur increasing naming for holistic composites, but decreasing it for feature composites. ANOVA was significant for the interaction between presentation (veridical and low blur) and system (holistic and feature) [$F(1,18) = 7.2$, $p = .015$, $\eta_p^2 = .29$]. A Simple Effects analysis reveals, from veridical to low blur, a reliable increase for holistic composites ($d = 0.68$), but a reliable and almost equal decrease for feature composites ($d = 0.66$) [2]. Clearly, even very mild levels of blur can have a substantial impact on naming that differs for different types of composite system.

---

[2] None of four contrasts involved here were significant ($p > .05$), even when appropriately pooling the (within- and between-subjects) error term to compare the two (between-subjects) system contrasts for veridical and low blur. A sensible conclusion is that the two slopes from veridical to low blur, for feature and holistic composites, are significantly different to each other.

Finally, we considered the impact of high blur. In this condition, feature detail is minimal and composite naming should be largely carried out by configural relations. Relative to veridical, high blur did not differ significantly by naming ($F < 1$) and system x presentation was not reliable ($F <$ 1). However, there was a significant benefit for naming of feature over holistic composites [$F(1,18) =$ 7.2, $p = .015$, $\eta_p^2 = .29$]. This advantage seems to have emerged from increased statistical power: naming was similar for veridical and high-blur composites, increasing the power of the system factor in the analysis and leading to the observed result. Since veridical and high blur are equivalent and have fairly-good naming ($M = {\sim}25\%$), the important implication is that veridical composites from both types of system have measurably useful configural relations for recognition.

For blurred composites shown in the second block, presentation x system was reliable [$F(2,36) = 7.5$, $p = .05$, $\eta_p^2 = .15$]. Simple Effects revealed an advantage of low-blur feature composites: (a) to low-blur holistic composites ($p < .001$, $d = 1.3$) and (b) to medium-blur feature composites ($p = .016$, $d = 1.0$). For feature composites presented second under low blur, this by-system advantage appears to have emerged due to participants having already seen holistic composites (also under low blur) in the first block. This carry-over effect (holistic to feature) is opposite to that found above for rotation (from feature to holistic); we consider these two differing effects in the General Discussion.

2.2.6. Stretch. Mean conditional naming scores were unexpectedly much higher for stretch relative to veridical for composites from both types of system. ANOVA was 2 (presentation: veridical and stretch) x 2 (image: composite and photograph) x 2 (order: first and second) x 2 (system: holistic and feature). This was significant for both presentation [$F(1,19) = 5.3$, $p = .033$, $\eta_p^2 = .22$] and image x presentation [$F(1,19) = 4.8$, $p = .042$, $\eta_p^2 = .20$]. Simple Effects indicated that stretch led to a reliable advantage for naming of composites ($p = .025$, $d = 0.55$) but not target photographs ($p = .43$).

We next asked the question whether stretch was more effective in the vertical or horizontal direction. ANOVA was 2 (direction: vertical and horizontal) x 2 (order: first and second) x 2 (system: holistic and feature) x 2 (image: composite and photograph). All main effects and interactions were non-significant ($F_s < 3.3$) except for direction x order x image [$F(1,18) = 4.6$, $p = .045$, $\eta_p^2 = .20$]. This three-way interaction indicated that there were no reliable differences in direction of stretch except for composites presented in the second block: then, naming was higher for vertical ($M = 50\%$) than horizontal ($M = 30\%$, $p = .049$, $d = 0.58$). As for negation, image stretch is likely to be an unusual combination for participants, and here it would seem that a block of vertically-stretched composites were required to be seen for effective use of the technique (or less effective, for horizontal).

We anticipated that the lack of naming benefit for stretched photographs may simply reflect ceiling-level performance for this type of image, as it is obviously not possible to increase performance beyond 100% correct. In this case, 15 of the 20 items were named perfectly in the target-familiarity check; for the remaining five, mean naming increased from 86.3% (SD = 7.2%) for veridical photos to 100% with stretching, a reliable improvement [$t(4)$ = 4.2, p = .013, d = 1.9].

To summarize, image stretch yielded a reliable naming advantage for composites, and for veridical photographs that had not been named initially with perfect accuracy. The data favoured linear stretch in the vertical over the horizontal direction.

2.2.7. Linear-regression analysis. In this section, the relative importance of featural and configural information were further assessed on conditional naming of (i) veridical composites, (ii) composites presented by negation, of which there is some debate about which information remains intact for photographs seen in this way (e.g. Bruce & Langton, 1994; Hole et al., 1999; Kemp et al., 1996) and (iii) stretched composites, of which we seem to recognise using configural relations (Hole et al., 2002). This assessment was achieved using linear regression, a standard statistical technique for indicating factors which exert a causal effect on the Dependent Variable (DV) (e.g. Field, 2005; Frowd et al., 2008).

Three regression analyses were conducted, each with predictors of 180° rotation and high blur. Recall that analyses resulting in a significant predictor for 180° rotation (inversion) generally indicate the importance of featural information for recognition, while those for high blur indicate the importance of configural cues. Two dichotomous predictors were included, system (1 = holistic and 2 = feature) and familiarity (0 = unfamiliar and 1 = familiar), the latter to explore whether composites were more identifiable when constructed of a familiar than an unfamiliar target (which research suggests that they should be). As the above results indicate that block order interacts with system and presentation, and for brevity, the current analyses involved (conditional) naming data collected from the first block. A backward model was run (probably of removal, p > .1), to indicate presence of suppressor variables: variables that exert a negative influence in the presence of other variables. Multicollinearity was not an issue here as the two non-dichotomous predictors (high blur and 180° rotation) were not highly correlated with each other (r = .16).

For the first regression model, the DV was veridical composite naming. The model achieved a good fit [$F(3,16)$ = 10.2, p = .001, R^2 = .66, Durbin-Watson dw = 2.1] with all predictors except for system emerging as reliable. Semi-partial (r_p) correlations were positive and similar in magnitude for

both high blur (B = 0.79, *SE(B)* = 0.19, r_p = .60) and 180° rotation (B = 0.72, *SE(B)* = 0.23, r_p = .47), and indicates that increases in accuracy in both of these components led to similar increases in naming of veridical composites. Familiarity was also positive (B = 1.2, *SE(B)* = 0.68, r_p = .26), indicating that more identifiable composites emerged when they were created from a familiar than an unfamiliar target. So, the results underscore the importance of both feature and configural information in veridical composites at naming, and the benefit of being familiar with a target at construction.

The other two regression analyses converged with (i) blur [B = 0.49, *SE(B)* = 0.12, r_p = .70] as a strong positive predictor for negation as DV [$F(1,18)$ = 11.1, p = .001, R^2 = .49, *dw* = 1.8], and (ii) both high blur [B = 0.84, *SE(B)* = 0.22, r_p = .68] and familiarity [B = 1.8, *SE(B)* = 0.77, r_p = .42] for stretch as DV [$F(2,17)$ = 10.1, p = .005, R^2 = .46, *dw* = 2.8]. These results indicate that configural information is involved when naming composites that are presented negated or stretched.

The same three analyses run using conditional naming scores of target photographs did not indicate the importance of either feature or relational information (Fs < 2.3). This null result is likely to have been caused by near-perfect naming of these stimuli, limiting experimental power.

2.3. Discussion

Experiment 1 involved eight main manipulations carried out on a set of composites. Results were detailed, especially for treatments involving different levels of blur but, in general, negation, rotation and blur reduced correct naming of composites, while stretch improved it. There were also different carry-over effects for rotation and blur by block order of presentation. We consider the impact of these effects in the General Discussion but, before then, follow-up on the unexpected, potentially-valuable forensic benefit of linear stretch. Recall that, relative to veridical, correct naming increased for participants inspecting composites seen with 100% vertical or horizontal stretch.

3.   Experiment 2: Perceptual size and perceptual stretch

A consequence of the previous experiment is that image stretch, if used on a composite in a criminal investigation, should substantially improve the chance that someone, typically a police officer or member of the public, will recognise the face and thereby assist the police to locate the offender.

It occurred to us that there is a pragmatic issue with this image format: composites look somewhat strange when stretched for the serious application for which they are normally used, to

catch serious criminals such as rapists and violent burglars. As such, police and media outlets may be reluctant to publish images shown this way. We thought that a more-acceptable solution would be for observers to look at an unaltered composite in such a way to give the appearance of stretch: viewing the face from the side or from underneath does just this. The outcome is a *perceptual* rather than a *physical* stretch.

We tested this seemingly-novel idea by asking participants to look at the face sideways, to produce the appearance of a vertical stretch, rather than the opposite, from below. This decision was based partly on Hole et al. (2002), who found that horizontal stretch can sometimes interfere with recognition of photographs, and also from the results of Experiment 1 here, which indicated a reliable benefit of vertical over horizontal stretch for composites presented in the second block.

An initial study was conducted using composites from Experiment 1, to see whether the technique would work in principle. Participants were given a printed copy of the composite and asked to look at the face sideways, to change viewpoint, and attempt to name the face. This perceptual-stretch condition led to reliably higher correct naming than other participants who were asked to do the same but were not given the instruction to change viewpoint. In the current experiment, we present a replication plus extension of this initial study.

We acknowledge that we have yet to provide an account of the stretch advantage, but one obvious explanation is that it is simply more engaging for an observer to look at a stretched than an unaltered face. A third condition was included to test for this possibility, with participants requested to move the composite towards and away from them, to allow the face to be seen at different sizes. This technique aimed to increase user engagement, but was not expected to alter recognition relative to veridical—that is, changing the visual angle in this way should not render a composite any more or less recognisable. We refer to this condition as *perceptual size*.

3.1. Method

3.1.1. Participants

Participants were an opportunity sample of 36 staff and student volunteers from the University of Central Lancashire. They were recruited on the basis of being familiar with celebrities and were different to participants who took part in the previous experiment (and the initial study). There were 17 males and 19 females and their age ranged from 18 to 42 (M = 27.0, SD = 6.4) years. Twelve people were assigned to each of the three levels of between-subjects factor, instruction type.

3.1.2. Design, Materials and Procedure

Materials were 15 printed EvoFIT composites created in a recent project (Frowd et al., 2012a, Experiment 3). Their construction was similar to that carried out for Experiment 1: participants looked at one of 15 familiar celebrities and then constructed a single composite of the face using EvoFIT. The composites were named with a mean of 24% correct.

Our participants were given one of three types of instruction, randomly assigned with equal assignment. In the control (normal-viewing) condition, they were asked to name the composites, guess if unsure or give a "no name" response. In the two other conditions, for different participants, they were then told to look at the face in a specific way as this may help recognition. One group was given the perceptual-stretch instruction—to look at the face sideways, to change viewpoint—while the other was given the instruction for perceptual size—to move the face towards and away from them. The researcher checked that the face was viewed front-on in normal-viewing and perceptual-size conditions, and side-on in perceptual stretch. Type of instruction was therefore a between-subjects factor (normal, perceptual stretch and perceptual size).

With the exception that participants were given different instructions for looking at the face for two out of the three conditions, and that composites were 15 veridical EvoFITs, the spontaneous-naming procedure was basically the same as that used in Experiment 1. Composites in the control condition were placed sequentially on a table in front of participants to keep the viewing distance roughly constant, but in the other two conditions, participants were handed the composites individually to allow them to view the face as described above. After naming the composites, participants were asked to name the target pictures, also presented sequentially. Stimuli were presented in a different random order for each person.

3.2. Results

As for the previous experiment, conditional naming scores were computed by dividing the number of correct names offered for each composite by the number of correct responses for the relevant target photograph. These data are summarized in Table 5. Participants moving the image backwards and forwards (perceptual image size) were slightly better at correctly naming the face than the standard presentation (normal viewing), but looking at the face from the side (perceptual stretch) elicited markedly higher correct naming.

Conditional naming scores were analysed by RM ANOVA. This by-items analysis was significant for instruction [$F(2,28) = 16.3$, $p < .001$, $\eta_p^2 = .54$] and simple contrasts indicated that perceptual image size and normal viewing were equivalent ($p = .31$), but perceptual stretch was superior to normal viewing ($p < .001$, $d = 0.77$); perceptual stretch was also superior to perceptual image size [$t(14) = 4.1$, $p = .001$, $d = 0.60$].

Table 5 about here

3.3. Discussion

Participants who looked at the face sideways, in the perceptual-stretch condition, correctly named the composite faces significantly better than participants who looked at the face front-on, the normal viewing condition. The effect size was large, indicating appreciable benefit of the method. The experiment replicates the initial study (described at the start of this experiment) and the vertical-stretch manipulation in Experiment 1. A reliable naming benefit therefore exists for composites that have been physically or perceptually stretched. In Experiment 2, perceptual stretch also led to significantly higher correct naming than perceptual size, indicating that the benefit of stretch is unlikely to be caused by increased task engagement.

4. General Discussion

Understanding how accurately people construct facial composites is theoretically interesting as well as providing potential avenues for improving the effectiveness of these forensic images. In the current project, we explored the extent to which featural and configural information was used for recognition of composites from two contrasting production methods: feature (E-FIT and PRO-fit) and holistic (EvoFIT) systems. In the first experiment, participants were given composites to name in one of eight main presentation types: veridical composites, or these same images negated, rotated (by 90° or 180°), with one of three levels of Gaussian blur (low, medium or high) or with 100% linear stretch. Participants first inspected images from one system and then the other. After naming the composites, they named photographs of the celebrities first in the same presentation as the composites, and then as a veridical image, to check their level of familiarity with the relevant target identities. A second experiment followed-up on the linear stretch manipulation.

4.1. Gaussian blur and image rotation.

It was hypothesised that feature composites would be more accurate in terms of individual features than configural information, spacing of features on the face, while holistic composites would show the opposite effect, with configural information being relatively more accurate than featural information. The correct naming data from Experiment 1 for Gaussian (blur) filtering provided some evidence to support this hypothesis—although, this was not quite as expected. For the mildest level of blur we used, only very fine details are concealed. However, this filtering influenced naming: for composites seen in the first block, recognition of holistic composites was greatly facilitated by low blur, while naming of feature composites was inhibited to the same extent. For example, the (veridical) feature composite of Mick Jagger in Figure 1 was named at 67% correct, but this figure dropped to 17% when seen with low blur. It would appear that very-detailed information is more-accurately constructed in feature than holistic composites: concealing the former (with low blur) reduces naming while concealing the latter improves it. In contrast, as naming was still possible under high blur, and to the level of veridical naming, this suggests that both feature and holistic composites contain configural relations that are useful for recognition to some extent.

A significant interaction was also observed between system and category of blur. For holistic composites presented in the first block, there was a significant linear (but not quadratic) decline in correct naming with increasing blur—from low to medium to high. This result ran contrary to expectation: increasing blur was not expected to influence naming for this type of composite. So, progressively concealing featural information led to worse correct naming. This pattern of decreasing naming with increasing blur matched naming of target photographs, a result that has been reported elsewhere using celebrity photographs (e.g. Lander et al., 2001). With the exception that holistic composites are deficient in high-detailed information (as indicated by low blur), holistic composites and photographs of faces are processed similarly with increasing levels of blur.

Interference to face recognition has been found to increase linearly with angle of rotation in the picture plane (e.g. Valentine & Bruce, 1988); also, that rotation tends to affect processing of configural relations, arguably more so for vertical, non-local distances between features (e.g. Rossion, 2008) and for the lower facial region (Tanaka & Gordon, 2011). In Experiment 1, interference was observed for both composites and target faces: an increase in rotation, from veridical to 90° to 180°, led to a linear decrease in correct naming. This transformation also revealed that, for holistic composites, those seen in the second block were better named than those seen in the first block. Inspecting rotated feature composites (first block)—images with somewhat accurate features (at least in terms of high-detailed information)—would appear to be leading to a processing

bias: a carry-over to naming of holistic composites. The opposite effect was observed for blurring at the mildest (low blur) level: holistic composites seen in the first block led to greater holistic processing of feature composites in the second block, and a naming advantage. Overall, the data indicate that both types of composite are somewhat accurate by both features and configural cues, information that can be disrupted by rotation and image blur.

In a linear-regression analysis, further evidence was sought for the role of featural and configural information. The approach used the strongest applications of rotation (180°) and blur (high) as predictors of veridical composite naming. High blur and 180° rotation emerged as reliable predictors (of composite naming) with positive part correlations; system was not reliable. These results support the above analyses which highlight the involvement of featural and configural information in both types of composite. A fourth predictor was included to indicate whether more-identifiable images were produced from face constructors who were familiar with the relevant identity. There was evidence that this predictor was also involved with a positive part correlation, implying that construction is more successful for familiar than unfamiliar faces. The result supports previous research for both feature (e.g. Davies et al., 2000; Frowd et al., 2011; Wogalter & Marwitz, 1991) and holistic composite systems (Frowd et al., 2011). Simply, it would appear that we have better access to the memory of a familiar than an unfamiliar face, and so construct a more-identifiable image of it.

4.2. Photographic negation.

Veridical composites and target photographs were correctly named less often in photographic negative than positive. This result for photographs is established (e.g. Bruce & Langton, 1994; Kemp et al., 1990; Phillips, 1971). For composites, the same as for photographs (e.g. Russell et al., 2006; Experiment 1), negation reduces but does not render recognition impossible since correct naming remained fairly good: here, it was 19.6% for negated composites. Also, naming increased for composites (but not targets) seen in the second than the first block. As mentioned earlier, this particular result is a likely consequence of two unusual formats, composites and negation, a novel situation that appears to become easier when more images are seen—in this case, in the second naming block (irrespective of whether these images were holistic or feature composites seen first).

There are several explanations as to what information is disrupted with photographic negation. These include disruption to pigmentation (e.g. Bruce & Langton, 1994), shape-from-shading relations (e.g. Johnston et al., 1992) and configural cues (e.g. Kemp et al., 1990). Our

regression analysis indicated that high blur was a positive predictor for naming of composites in photographic negative; in other words, both negation and high blur have a tendency to reveal similar information, in this case configural relations. These data provide evidence against our prediction that negation interferes with configural cues. In contrast, it appears that facial features are disrupted to a greater extent, thus supporting Bruce and Langton's (1994) work. It is perhaps difficult to process composite images given disturbance to the appearance of skin and hair. Such a suggestion makes good sense given recent research (to be published) indicating the surprising and sizeable negative influence on recognition for composites with altered hair.

4.3. Linear stretch.

Hole et al. (2002) found that linear, vertical stretching of photographs of faces did not influence recognition. They also found that photographs of faces remained recognisable when stretched and blurred, highlighting the importance of configural cues for recognising a stretched image. Here, correct naming of stretched composites reliably increased (Experiment 1); the effect was found to be independent of composite system and block order, indicating consistent effects. The regression analysis revealed that high blur was a reliable predictor for naming of stretched composites and suggests, as for negation, that participants were recognising such images from configural cues.

In Experiment 1, there was also an interaction between image and presentation type which indicated that the overall benefit of image stretch was restricted to composites. The majority of veridical photographs (15 out of 20 items) were named at 100% and thus any benefit of stretch was not possible to measure. This is a typical result for unaltered photographs: ceiling-level recognition. As a result, some studies deliberately reduce recognition by blurring, using brief presentation of the stimuli, or by measuring reaction time as the DV (e.g. Lander et al., 2001; Lee & Perrett, 2000; Valentine & Bruce, 1988). However, five of our veridical photographs were named less than perfectly and, for these items, correct naming reliably increased when seen stretched. To our knowledge, this is the first study to demonstrate a recognition benefit of linear stretch for photographs of faces, a result that is worthy of replication. Currently, it is not known which level of stretch gives best results, although this must be bounded: clearly, any image will eventually lose all information if stretched too far in either the vertical or horizontal direction.

Vertical stretch promoted reliably better naming than horizontal stretch for composites viewed in the second block. This indicates that the vertical format is more effective than the horizontal. This idea broadly fits with the finding of Hole et al. (2002) who found that horizontal

stretch can inhibit recognition for photographic faces. While the reason for this is not entirely clear, one possibility is that configural cues are simply easier to extract in the vertical case as the entire image can be seen more easily: wide (horizontally stretched) faces may require a visual scan to perceive them in their entirety. This idea can be tested by varying visual angle: if our explanation is correct, then naming should improve with large relative to small visual angles for horizontally- but not vertically-stretched composites.

A replication of the vertical stretch advantage for composites was attempted in Experiment 2. It was thought that police would be reluctant to publish stretched images in the media, as they do not look sensible (appropriate) for their forensic role. Our solution was to maintain the pictorial aspect of the composite but ask observers to look at the face from the side—although, there are other ways to legitimise stretch (e.g. http://tiny.cc/pbi-composite). The benefit of this *perceptual* stretch successfully extended to a different set of composites and was also better than a third condition which aimed to increase user engagement, by varying the composites' apparent size.

As mentioned earlier, the perceptual-stretch benefit has also been replicated elsewhere. In Frowd et al. (2013), participants looked at a video containing an unfamiliar target face and, 24 hours later, received one of two types of interview and constructed a composite. While the interview for one group of participants involved freely recalling the target face, as was the case here, the other group did the same but then focussed on the character (personality) of the target face. Naming of the resulting composites was higher when looking at the face from the side than the front; however, perceptual stretch was particularly effective for composites constructed after the character-type interview, a technique that is known to produce an overall more identifiable image (e.g. Frowd et al., 2008). These data suggest that image stretch is likely to involve processing of the face as a whole, a result that is supported by our data as well as Hole et al. (2002). In fact, our regression analysis indicated that naming of stretched composites was predicted by both high blur and familiarity with the target at construction. While high blur indicates holistic (configural) face processing, a change in familiarity, from unfamiliar to familiar, seems to involve a qualitative shift in processing from more component-based to whole-face recognition (e.g. Megraya & Burton, 2006).

Other approaches have successfully enhanced recognition of finished composites. Bruce et al. (2002) found that combining either two or four feature composites of the same identity produced an average image that was more effective at identification than any of the constituent images on average. The authors explained that as composites had been constructed independently of each other, errors present in the faces would be uncorrelated and so would cancel during averaging. Recent research indicates that this effect extends to holistic composites (Valentine et al., 2010).

Similarly, Burton, Jenkins, Hancock and White (2005) found that averaging photographs of faces of the same identity produces a representation that is effective for both human and computer face-recognition systems. For photographs and composites then, variability in the resulting representation is reduced: the average tends to look overall more similar to the relevant person.

The stretch techniques may similarly reduce within-person variation, but the mechanism by which this is achieved is likely to be different. We do not normally perceive faces in a stretched format and yet this representation (particularly in the vertical direction) does not interfere with recognition of photographs of faces (e.g. Hole et al., 2002). Hole et al. (2002) argue that a stretched face is normalised (made more 'face like') as part of cognitive processing to match our veridical memories (cf. deformable template theories). It is likely that this process reduces error, in this case by making a composite face more similar to familiar identities. Providing that the relational information is somewhat accurate in a composite (esp. after a character-type interview or with face construction of a familiar identity), recognition would be facilitated by an increase in accuracy of facial features. On-going research is exploring this issue.

So, forensic practitioners are likely to achieve best performance using the character-type interview (http://tiny.cc/holistic-ci) and one of the stretched procedures for composite naming (e.g., http://tiny.cc/pbi-composite). We suggest that composites are accompanied by an instruction for how to view the image: "viewing the composite sideways, to give a different perspective, may help you to recognise the face". The technique has been trialled in UK policework and is in current use.

4.4. Additional practical implications.

The current design involved composites from feature and holistic systems that were selected to be equally identifiable overall, allowing an assessment of the relative efficacy of featural and configural cues for recognition. Overall, both systems appear to contain featural and configural information that are useful for recognition, but arguably the major difference between them relates to highly-detailed information: low blur indicated that fine details tended to be accurate in feature composites but inaccurate in holistic composites.

It is likely that system differences were found because of the way in which the composites were constructed. For E-FIT and PRO-fit, constructors made selections of facial features. These facial elements were switched in and out of an intact face and their task was to decide which items created the best likeness. This method, by seeing features change sequentially in a face, is likely to result in constructors discriminating fine detail: skin and texture, exact outline of each feature, detail

of hair, etc. For EvoFIT, constructors were asked to select items based on overall appearance, to evolve a face, and then manipulate the face to improve holistic aspects—age, masculinity, weight, etc. Using this method, the focus is not on fine detail. It would be interesting now to see whether people could construct EvoFIT images with more-accurate detail by prompting them to take this into account when making selections. Ultimately, all constructors were given the opportunity to position individual features, which is arguably why configural cues were found to be valuable for composites not just from EvoFIT but also from the feature systems.

These observations hint at why feature composites do not enjoy good recognition when constructed after a long retention interval. After a couple of days, as is normal in police practice, such images are typically named with a mean of around 5% correct (e.g. Frowd et al., 2005a, 2007b, 2007c, 2010). People struggle to recall details of the face, and recall becomes worse with delay (Ellis, Shepherd & Davies, 1980), but such information is likely to be important. If it is the case that featural cues are erroneous in composites constructed after long delays, one solution might be to publish composites in the media with a high level of blur. Given the large number of police practitioners who use technology of this type, particularly in the USA (McQuiston-Surrett, Topp & Malpass, 2006), resolving this issue would appear to be worthwhile.

The designers of EvoFIT aimed to produce a recognisable image of an unfamiliar face following a long retention interval (e.g. Frowd et al., 2004). For a delay of 24 hours, research using the latest version of EvoFIT indicates that its composites are named (front-on) at around 50% correct (Frowd et al., 2012b, 2013). Based on the current results, even-higher performance may be achieved if facial detail could be more-accurately rendered—although, it is unclear whether constructors are capable of doing this. Clearly, an alternative is to simply conceal this seemingly inaccurate information using mild blur. However, this approach may not be necessary: the most-recent evaluation of EvoFIT (Frowd et al., 2013) described above reveals that an extraordinarily-high naming level of 74% correct is now possible using a combination of techniques that (a) enhance holistic processing at face construction (Frowd et al., 2008, 2012b), (b) employ the latest EvoFIT face-construction software (Frowd et al., 2012b) and (c) involve stretch at naming.

Other holistic systems exist. These include EFIT-V and ID (Gibson et al., 2005; Tredoux et al., 2006), and their method of constructing a face is similar to that used with EvoFIT: witnesses repeatedly select from arrays of complete faces. At least one of them, EFIT-V, has holistic tools to enhance an evolved face, along with other tools to manipulate size and placement of individual features. There is evidence that EFIT-V produces composites that have fairly-good naming, at least following a very-short retention interval (Valentine et al., 2010), and so one would anticipate that

the current results would extend to other holistic systems that are tested in the same way. Similarly, given the fairly-common approach used to construct a face for feature systems, one would imagine that our findings would also extend to them (at least when construction delays are fairly short).

4.5. Summary.

The current project sought to understand the effectiveness of featural and configural information in composites constructed from feature and holistic systems. Both types of information were found to be valuable in both types of composite, although feature composites (PRO-fit and E-FIT) were more accurate than holistic (EvoFIT) composites in terms of highly detailed information. It would appear that techniques which improve (or conceal) this information may provide a useful enhancement to practitioners who use systems after long retention intervals; and, different levels of image blur may achieve this objective (although this may not be necessary for EvoFIT due to high naming). Also, looking at a composite that had been linearly stretched greatly increased an observer's ability to correctly name the face. A practical extension of this technique, with observers looking at a veridical face sideways, similarly facilitated recognition. This makes for a technique that, if continued to be used in police investigations, has the potential to substantially increase identification of suspects.

References

Benson, P. J., & Perrett, D. I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology, 3*, 105-135.

Bindemann, M., Burton, A. M., Leuthold, H., &Schweinberger, S. R. (2008). Brain potential correlates of face recognition: Geometric distortions and the N250r brain response to stimulus repetitions. *Psychophysiology*, 45, 535–544.

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.

Bruce, V., Burton, A.M., & Craw, I.G.C. (1992). Modelling face recognition. *Philosophical Transactions of the Royal Society of London, Series B*, 335, 121-128.

Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. (1992). The importance of "mass" in line-drawings of faces. *Applied Cognitive Psychology*, 6, 619- 628.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognizing the sex and identities of faces. *Perception*, 23, 803–822.

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.

Bruce, V., Valentine, T., & Baddeley, A.D. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, 1, 109-120.

Bruyer, R., Galvez, C., & Prairial, C. (1993). Effect of disorientation on visual analysis, familiarity decision and semantic decision on faces. *British Journal of Psychology*, 84, 433-441.

Burton, A.M., Jenkins, R., Hancock, P.J.B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 51, 256-284.

Burton, A.M., Jenkins, R., Hancock, P.J.B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 51, 256-284.

Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science*, 11, 429-433.

Carbon, C., & Leder, H. (2005). When feature information comes first! Early processing of inverted faces. *Perception*, 34, 1117-1134.

Collishaw, S.M., & Hole, G.J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception,* 29, 893-909.

Collishaw, S.M., & Hole, G.J. (2002). Is there a linear or a nonlinear relationship between rotation and configural processing of faces? *Perception*, 31, 287-296.

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.

Dakin, S.C., & Watt, R.J. (2009). Biological "bar codes" in human faces. *Journal of Vision*, 9, 1–10.

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

Davies, G.M., Ellis, H.D., & Shepherd, J.W. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, 92, 507-523.

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.

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

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.

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.

Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. R. (1998). What is "special" about face perception? *Psychological Review*, 105, 482-498.

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.

Field, A. (2005). *Discovering statistics using SPSS*. Second edition. Sage: London.

Freire, A., Lee, K., & Symons, L.A. (2000). The face-inversion effect as a deficit in the encoding of configural information: direct evidence. *Perception*, 29, 159–170.

Frowd, C.D., Bruce, V., & Hancock, P.J.B. (2009). Evolving facial composite systems. *Forensic Update*, 98, 25-32.

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

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

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

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

Frowd, C.D., Herold, K., McDougall, M., Duckworth, L., Hassan, A., Riley, A., Butt, A., McCrae, D., Wilkinson, C., Skelton F.C., McIntyre, A., & Hancock, P.J.B. (2012a). Hair today, gone tomorrow: holistic processing of facial-composite images. Manuscript under revision.

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

Frowd, C.D., Nelson, L., Skelton F.C., Noyce, R., Atkins, R., Heard, P., Morgan, D., Fields, S., Henry, J., McIntyre, A., & Hancock, P.J.B. (2012b). Interviewing techniques for Darwinian facial composite systems. *Applied Cognitive Psychology,* DOI: 10.1002/acp.2829.

Frowd, C.D., Petkovic, A., Nawaz, K., & Bashir, Y. (2009). Automating the processes involved in facial composite production. In A. Stoica, T. Arslan, D. Howard, T. Higuchi, and A. El-Rayis (Eds.) *2009 ECSIS Symposium on Bio-inspired, Learning, and Intelligent Systems for Security* (pp. 35-42). NJ: CPS.

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

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

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

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

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

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science*, 19, 207-208.

George, N., Dolan, R.J., Fink, G.R., Baylis, G.C., Russell, C., & Driver, J. (1999). Contrast polarity and face recognition in the human gyrus. *Neuroscience*, 2, 574-580.

George, P.A., & Hole, G.J. (2000). The role of spatial and surface cues in the age-processing of unfamiliar faces. *Visual Cognition*, 7, 485-509.

Gibson, S.J., Solomon, C.J., & Pallares-Bejarano, A. (2005). Nonlinear, near photo-realistic caricatures using a parametric facial appearance model. *Behavior Research Methods,* 37, 170-181.

Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and performance*, 33, 995-1002.

Hancock, P.J.B., Bruce, V., & Burton, A.M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4-9, 330-337.

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.

Hole, G.J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, 23, 65-74.

Hole, G.J. (2011). Identity-specific face adaptation effects: Evidence for abstractive face representations. *Cognition*, 119, 216-228.

Hole, G.J., George, P.A., & Dunsmore, V. (1999). Evidence for holistic processing of faces viewed as photographic negatives. *Perception*, 28, 341-359.

Hole, G.J., George, P.A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, 31, 1221-1240.

Jacques, C., & Rossion, B. (2007). Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *NeuroImage*, 36, 863–876.

Johnston, R.A., & Edmonds, A.J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, 17, 577-596.

Johnston, A., Hill, H., & Carman, N. (1992). Recognising faces: effects of lighting direction, inversion, and brightness reversal. *Perception*, 21, 365-375.

Johnston, A., & Passmore, P.J. (1994). Shape from shading. I: Surface curvature and orientation. *Perception*, 23, 169-189.

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. *Perception*, 19, 531-543.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues. *Perception*, 25, 37-52.

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.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology*, 51A, 449-473.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: the role of configural information in face recognition. *Quarterly Journal of Experimental Psychology: Section A,* 53, 513-536.

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.

Lewis, M.B. (2001). The lady's not for turning: Rotation of the Thatcher illusion. *Perception*, 30, 769-774.

Lewis, M.B. (2006). Eye-witnesses should not do cryptic crosswords prior to identity parades. *Perception*, 35, 1433-1436.

Liu, C.H., Collin, C.A., Burton, A.M., & Chaurdhuri, A. (1999). Lighting direction affects recognition of untextured faces in photographic positive and negative. *Vision Research*, 39, 4003-4009.

McQuiston-Surrett, D., Topp, L. D., & Malpass, R. S. (2006). Use of facial composite systems in U.S. law enforcement agencies. *Psychology, Crime and Law*, 12, 505–517.

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

Moses, Y., Ullman, S., & Edelman, S. (1996). Generalization to novel images in upright and inverted faces. *Perception*, 25, 443-462.

Murray, J. E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science*, 11, 492-496.

Patterson, K.E., & Baddedley, A.D. (1977). When recognition fails. *Journal of Applied Psychology*, 3, 406-417.

Phillips, R.J. (1972). Why are faces hard to recognize in photographic negative? *Perception, & Psychophysics*, 12, 425-426.Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, 128, 274-289.

Rossion, B., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. *Journal of Vision*, 8, 1-13.

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception*, 35, 749-759.

Schwaninger, A., Lobmaier, J. S., & Collishaw, S. M. (2002). Role of Featural and Configural Information in Familiar and Unfamiliar Face Recognition. *Lecture notes in computer science*, 2525, 643-650.

Schooler, J.W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology*, *16*, 989-997.

Schwaninger, A., & Mast, F.W. (2005).The face-inversion effect can be explained by the capacity limitations of an orientation normalization mechanism. *Japanese Psychological Research*, 47, 216-222.

Searcy, J.H., & Bartlett, J.C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 904-915.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, 75, 221-243.

Sinha P., & Poggio T. (1996). I think I know that face. *Nature*, 384, 404–404.

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

Tanaka, J.W., & Gordon, I. (2011). Features, Configuration, and Holistic Face Processing. In Calder, A.J. (Ed.). *The Oxford handbook of face perception* (pp. 177-194). New York: Oxford University Press.

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

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.

Troje, N., & Bülthoff, H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, 36, 1761-1771.

Valentine, T, & Bruce, V. (1988). Mental rotation of faces. *Memory & Cognition*, 16, 556-566.

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

Vuong, Q.C., Peissig, J.J., Harrison, M.C., & Tarr, M.J. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and Greebles. *Vision Research*, 45, 1213-1223.

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

White, M. (2001). Effect of photographic negation on matching the expressions and identities of faces. *Perception*, 30, 969-981.

Wogalter, M., & Marwitz, D. (1991). Face composite construction: In view and from memory quality improvement with practice. *Ergonomics*, 22, 333–343.

Yasuda, T. (2005). Effects of configural information on face shape perception. *Japanese Journal of Psychology*, 76, 131-138.

Yip, A.W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception*, 31, 995-1003.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configurational information in face perception. *Perception*, 16, 747-759.

List of tables

Table 1. Configural and featural information which research would suggest is principally used for recognising photographs of faces seen with different types of image transformation

Negation	Rotation	Blurring	Stretching
Featural	Featural and some aspects of configural	Configural	Configural

Table 2. Conditional naming scores (expressed in percentage-correct) in Experiment 1 for both celebrity composites and target photographs at eight levels of presentation (Veridical … Stretch). See accompanying text for calculation of this DV.

Image	Veridical	Negation	Rotation		Gaussian blur			Stretch
			90°	180°	Low	Medium	High	
Composites	29.7† (5.8)	19.6* (4.2)	24.0† (4.5)	18.8† (4.4)	27.2[1] (3.7)	16.6[1] (2.8)	21.8[1] (4.9)	41.6* (5.7)
Targets	97.8† (2.1)	88.0* (2.9)	89.0† (2.0)	82.4† (4.4)	97.5‡ (1.0)	89.7‡ (2.3)	79.8‡ (4.6)	99.6[2] (0.4)

Note. Values in parenthesis are standard errors of the by-item means. *Significantly different to Veridical, $p < .05$. †Significant linear trend (in the left-to-right order shown including Veridical), $p < .01$. ‡Significant linear trend (in the left-to-right order shown excluding Veridical), $p < .01$. [1]Significant interaction between order and system, $p < .001$. See text and Table 4 for analysis of this interaction. [2]Significantly different to Veridical for those identities ($N = 5$) that were not named at 100% in the target-familiarity check, $p < .02$.

Table 3. Percentage-correct naming illustrating the block order (first and second) x system (EvoFIT and feature) interaction for image rotation. Data are shown collapsed over image type (composites and target photographs).

	First	Second
System		
Holistic	51.8§ (4.4)	58.9§ (3.7)
Feaure	61.1 (2.9)	55.9 (4.7)

Note. Values in parenthesis are standard errors of the by-item means. §Significantly different to each other, $p < .05$.

Table 4. Percentage-correct naming for composites presented by block for increasing levels of categorical blur.

System	Veridical	Blur		
		Low	Medium	High
First block				
Holistic	16.7[1]	33.7[1]‡	20.8‡	10.0‡
	(7.9)	(8.0)	(5.4)	(6.7)
Feature	36.3[1]	20.3[1]	18.5	40.0
	(8.2)	(7.0)	(4.0)	(7.0)
Second block				
Holistic	27.0	14.0§	10.3	19.2
	(8.3)	(5.4)	(4.5)	(8.4)
Feature	38.8	40.7§	16.7§	18.0
	(11.5)	(8.0)	(6.6)	(6.8)

Note. Values in parenthesis are standard errors of the by-item means. [1]Significant system x presentation interaction, $p < .02$. ‡Significant linear trend within presentation type (in the left-to-right order shown excluding Veridical), $p < .05$. §Advantage for Feature composites under low-blur, $p < .05$.

Table 5. Percentage-correct naming for celebrity composites inspected in the normal way, at a fixed viewing distance (Normal viewing); with participants altering the viewing distance for themselves (Perceptual image size); and with participants looking at the face from the side (Perceptual stretch).

Normal viewing	Perceptual image size	Perceptual stretch
17.1	20.2	35.8**
(5.2)	(6.3)	(7.1)

Note. Values in parenthesis are standard errors of the by-item means. **Significantly different to the two other conditions, $p < .002$.
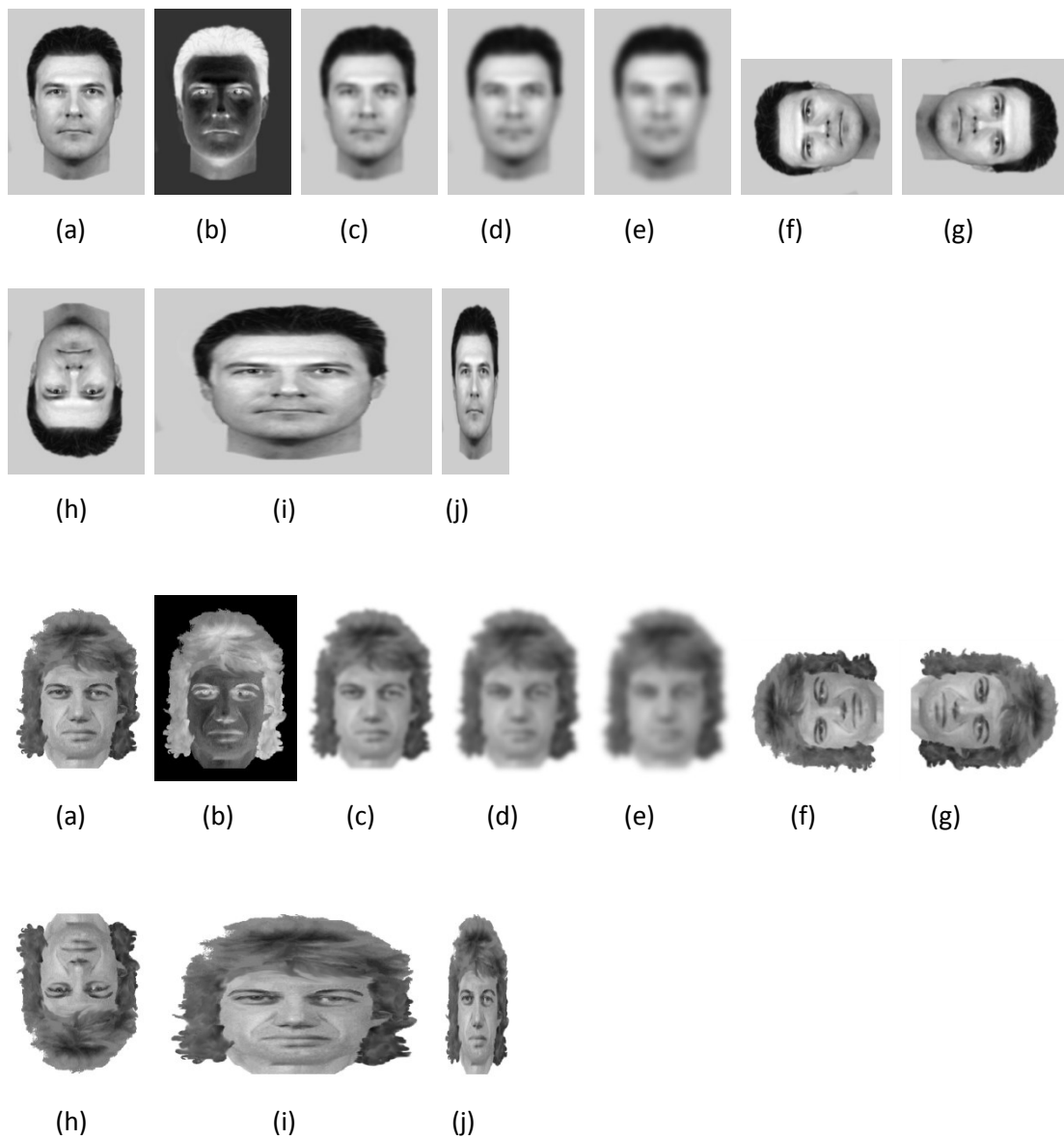
Figure 1. Example composites and presentation used in Experiment 1. They are a holistic (EvoFIT) composite of TV celebrity Simon Cowell (top two rows), and a feature (E-FIT) composite of popstar Mick Jagger (bottom two rows). In Experiment 1, participants were presented with 20 such composite images to name from one of the following: (a) veridical, (b) photographic negation, (c) low blur, (d) medium blur, (e) high blur, (f) 90° rotation to left, (g) 90° rotation to right, (h) 180° rotation (inversion), (i) stretched horizontally by 100% and (j) stretched vertically by 100%.