Understanding the multi-frame caricature advantage for recognising facial composites

Charlie D. Frowd (1*)
Faye Skelton (1)
Chris Atherton (1)
Melanie Pitchford (2)
Vicki Bruce (3)
Rebecca Atkins (1)
Carol Gannon (1)
David Ross (4)
Fern Young (4)
Laura Nelson (1)
Gemma Hepton (1)
Alex H. McIntyre (5)
Peter J.B. Hancock (5)

(1) School of Psychology
University of Central Lancashire, PR1 2HE
* Corresponding author: Charlie Frowd, School of Psychology, University of Central Lancashire, Preston PR1 2HE, UK. Email: cfrowd@uclan.ac.uk. Phone: (01772) 893439.
(2) Department of Psychology
Lancaster University, LA1 4YF
(3) School of Psychology
Newcastle University, NE1 7RU
(4) School of Psychology
Cardiff University, CF10 3AT
(5) Psychology, School of Natural Sciences
University of Stirling, FK9 4LA

Abstract

Eyewitnesses often construct a ‘composite’ face of a person they saw commit a crime, a picture that police use to identify suspects. We described a technique (Frowd et al., 2007, *Visual Cognition*, 15, 1-31) based on facial caricature to facilitate recognition of these images: correct naming substantially improves when composites are seen with progressive positive caricature, where distinctive information is enhanced, and then with progressive negative caricature, the opposite. Over the course of four experiments, the underpinnings of this mechanism are explored. Positive-caricature levels were found to be largely responsible for improving naming of composites, with some benefit from negative-caricature levels. Also, different frame-presentation orders (forward, reverse, random, repeated) facilitated equivalent naming benefit relative to static composites. Overall, the data indicate that composites are usually constructed as negative caricatures.

Running head: Multi-frame caricaturing

Keywords: facial composite, caricature, face space, memory, witness

Caricature is a technique that distorts the overall appearance of a face, exaggerating its unique characteristics, and its effect on human face perception has attracted considerable research interest (e.g. Benson & Perrett, 1991; Rhodes & Tremewan, 1994; Tanaka & Simon, 1996). Artists can produce extreme caricatures that remain recognisable, even from images containing only a few brush strokes, suggesting that such distortions retain psychologically-relevant information. In contrast, computer-generated facial caricatures produce more consistent results than those from artists (Benson & Perrett, 1991), and normally involve exaggerating distinctive ‘shape’ information—the outline of unusual facial features and the relational information, the distances between features.

There is evidence that face images transformed by caricature are recognised more quickly and/or more accurately than veridical images (see Rhodes, 1996, for a review). A positive level of (shape) caricature, in which the geometric characteristics of the face are exaggerated with respect to an average face, improves recognition of outline drawings (Benson & Perrett, 1994; Tanaka & Simon, 1996). This effect of positive caricature tends not to occur for photographic images of familiar faces due to our near-perfect ability to recognise people that we know well (e.g. Bruce, 1982); it does, however, produce more self-priming than veridical (or negatively-caricatured) familiar faces (Calder, Young, Benson & Perrett, 1996), as well as improving recognition of images that have been ‘degraded’—for example, photographs that are presented briefly (Lee, Byatt & Rhodes, 2000; Lee & Perrett, 1997) or presented as line drawings (e.g. Benson & Perrett, 1994; Rhodes & Tremewan, 1994; Tanaka & Simon, 1996). There is also evidence that positive caricature of facial texture (pixel intensity) facilitates recognition for briefly presented photographs (Lee & Perrett, 2000). Negative caricature, in which distinctive information is attenuated, reduces recognisability (e.g. Lee et al., 2000; Rhodes, Carey, Byatt & Proffitt, 1998).

These data can be explained in terms of two models of face space proposed by Valentine (1991). These models are norm-based and exemplar-based, and considerable research effort has been spent on deciding which type best captures the available data (e.g., Giese, Sigala, Wallraven, & Leopold, 2004; Jeffery et al., 2010; Lewis, 2004; Ross, Deroche & Palmeri, 2011); some authors also argue for the involvement of non-linear types (e.g., Gibson, Solomon & Pallares-Bejarano, 2005; Susilo, McKone, & Edwards, 2010). For an up-to-date review, see Rhodes and Leopold (2011).

In a traditional norm-based model, as illustrated in Figure 1, memories of familiar people are conceptualised as residing within a multi-dimensional face space, representations that are encoded by their direction of deviation and distance from an average or 'norm' face. There is evidence that faces which are average (typical) in appearance are clustered around this norm, while others that are more distinctive in appearance reside further away (see Johnston, Milne, Williams & Hosie, 1997, for empirical evidence). In this space, the angle between one memory and another with respect to the central norm as well as vector length allow for discrimination of identity—see Figure 3 for an example. Positive caricature artificially inflates facial distinctiveness by shifting representations along an imaginary trajectory away from the norm, increasing vector length, to a region of space with less competition from other memories. This transformation can facilitate recognition. For example, Benson and Perrett (1994) found overall more accurate recognition of line-drawings where the shape information in the faces had been positively caricatured. Conversely, faces that translate along this trajectory towards the norm, to give a negative or anti-caricature, reduce distinctiveness, increase competition and are recognised with lower accuracy.

Exemplar-based models are similarly conceptualised using multi-dimensional face space. They involve an absolute rather than a relative coding scheme. The example given in Figure 1, without axes necessarily being located centrally, would be such a model. For more details and for an example model of this type, refer to Lewis (2004).

Figure 1 about here

A different class of stimuli, facial composites, may also benefit from positive caricature. Facial composites are constructed by eyewitnesses of people they have seen commit a crime, and police use such images to locate these offenders. Composite images are traditionally constructed using sketch artists, with pencils or crayons, and ‘feature’ systems, by the selection of individual facial features (eyes, hair, nose, mouth, etc.). However, correct naming rates of such composites tend to be low (Brace, Pike & Kemp, 2000; Bruce, Ness, Hancock, Newman & Rarity, 2002; Davies, van der Willik & Morrison, 2000; Frowd et al., 2005b, 2007a, 2010; Koehn & Fisher, 1997), which is one of the reasons that attempts have been made to create alternative, ‘holistic’ systems whereby eyewitnesses select from arrays of complete faces (e.g. Frowd et al., 2012; Gibson, Solomon, Maylin & Clark, 2009; Valentine, Davis, Thorner, Solomon & Gibson, 2010). One possible reason for poor identification with traditional methods is that composites can appear quite similar to one another (Frowd et al., 2005a), and this similarity may lead to confusion about the faces’ identities, or even failure to recognise them at all. This is analogous to the problem of trying to identify one of the many typical faces clustered in the centre of the conceptual model shown in Figure 1.

In 2007, we (Frowd, Bruce, Ross, McIntyre & Hancock) explored the potential benefit to recognition of viewing caricatured composite images. To do this, we used the same basic procedure to caricature the shape aspects of the face as carried out in past research (e.g. Benson and Perrett, 1994). It was found that, for composites given a single fixed level of positive (shape) caricature, spontaneous naming rates did not improve overall. The reason appeared to be that, for a given composite, the preferred level of caricature varied considerably between observers. We argued that recognition would be facilitated if each observer viewed a range of caricature levels. Results confirmed this suggestion: mean naming was 28.8% correct when participants viewed veridical (static) composites, and this increased to 42.4% when they viewed static composite plus a range of caricature levels—images caricatured at 20 levels from -30% (negative caricature) and +30% (positive caricature). See Figure 2 for an example. We found that the benefit occurred for both computer-generated composites and sketched composites—although, due to a general preference for more extreme caricaturing in sketches, the effect was observed over a wider range, -50% to +50%.

Figure 2 about here

Four further results from Frowd et al. (2007b) are worth noting. First, exposure to different levels of caricature was required for a reliable benefit to be observed. This held true even for composites caricatured at an optimal level, images based on participants' average preferred level of exaggeration (which varied between faces). Second, benefit was found for composites constructed from famous as well as unfamiliar targets. Third, an advantage occurred for the 21 individual frames presented consecutively on A3 paper (a 'photospread'), or as an animation, an image cycling continuously over the same range every 6 seconds. Fourth, the caricaturing advantage was conferred by three different methods of face production from memory: artist sketches, two traditional 'feature' systems (PRO-fit and E-FIT), and the 'holistic' EvoFIT system. Overall, the technique demonstrated robustness by method of production and type of presentation.

The authors proposed two explanations for the multi-frame (photospread and animation) caricaturing advantage. One explanation was that a single level of positive caricature—one frame in that part of the sequence—was optimal for recognition, as this would present an exaggerated level of distinctiveness and trigger recognition. Support for this explanation was based on general findings in face recognition for the benefit of caricature (e.g. Benson & Perrett, 1991). The other explanation was based on the fact that composites are not perfect representations; they contain errors of feature shape, feature placement and greyscale shading. As such, the anti-caricaturing part may minimise such error, facilitating recognition. This notion was supported by their finding that observers tended to believe that anti-caricatured composites were better likenesses (e.g., M = -10% for feature composites). As they presented entire sequences, it was difficult to establish which explanation was correct.

The intention of the current work was to resolve which of these explanations are correct, or indeed whether both might play a part in the caricature advantage. We also wanted to explore the extent to which the multi-frame caricaturing benefit would extend to different task conditions (e.g., random frame order, presentation at forward, backward and reduced frame rate), and to provide a framework in which to understand the effect.

Given the apparent equivalence of norm and exemplar models, we propose a more abstract conceptualisation here with the aim of not invoking either type. Clearly, under such a generic "face space" account, multi-frame caricaturing progressively increases the level of distinctiveness starting from a location (an anti-caricature) close to the reference (norm) face used for caricaturing and along a trajectory (vector) that passes through a representation of the composite itself (veridical) and outwards with increasing positive caricature. The four experiments presented here establish which region or regions along this trajectory are responsible for the multi-frame caricature advantage.

It is perhaps worth emphasising a divergence from the face-recognition literature. As mentioned above, while positive caricature facilitates naming of line drawings, no such effect exists for positive caricature of photographic stimuli, except when recognition is rendered more difficult (e.g. under image degradation, brief presentation). On the other hand, composites are an imperfect representation and are rarely (if ever) recognised with perfect accuracy (e.g. Frowd et al., 2005b). It is for this reason that there is scope for a recognition benefit for composites presented at multiple levels of caricature.

Our point of departure is to explore identification of composites over separate positive and negative caricature ranges (Experiments 1 and 2). It is found that the positive-caricature range is largely responsible for the multi-frame naming advantage. Further evidence for the contribution of this caricature range was sought in Experiments 3 and 4, as well as (1) investigating the level of exaggeration at which recognition tends to occur, (2) whether the direction in which frames are presented (forward, reverse, random) is important, and (3) whether any benefit is conferred from sequence repetition.

Experiment 1: Effectiveness of negative and positive caricature sub-ranges

In the first experiment, we investigated the relative contributions of the negative and positive sections of Frowd et al.'s full-range animation to successful naming (±30% caricature range for 'feature' and 'holistic' composites, and ±50% for sketch). These two components of the animation have not been considered separately in previous naming tasks; here, their effectiveness was compared against viewing of full-range caricature and static composite. We chose multi-frame caricature presented as animated images rather than as a series of photospreads containing the same frames since the former is the preferred option for police use; there is also evidence (Frowd et al.) that these two formats promote equivalent recognition (this is also found here in Experiment 3).

Method

Design

The current experiment involved composites of well-known male celebrities. As detailed in the Materials, these images were produced from three different systems, and also with some produced by participants familiar with the face and others where the face was unfamiliar, as is more common forensically. Further, composites were produced after a delay of a few hours from seeing the target face or after a forensically-realistic 2-day delay. While differences such as these yield composites of different quality, Frowd et al. (2007b) found that multi-frame caricaturing was effective irrespective of system, target familiarity and retention interval. Note that in Experiment 2, we attempt a replication under more carefully-controlled face-construction procedures.

Participants were presented with 18 composites in one type of presentation (static / negative / positive / full range) and so the design was between-subjects for this factor.

Participants

Participants were 14 female and 42 male volunteer staff and students at the University of Central Lancashire (UCLan). Their ages ranged from 23 to 61 years (*M* = 41.7 years, *SD* = 9.0 years). They were recruited on the basis of being generally familiar with well-known male celebrities and (as elsewhere in the paper) were allocated equally to conditions in the between-subjects factor, presentation type.

Materials

The celebrity stimuli used by Frowd et al. (2007b) were employed. These were originally extracted from archives and comprised six sketches, six 'feature' composites, and six 'holistic' composites of well-known identities. To produce an image, participants viewed a photograph of a celebrity for 60 seconds, and then described it out loud, following which they constructed a single composite image. The procedure used to produce the images is detailed (see Frowd et al., 2007b, for a full description), but in brief: sketches were drawn by hand from feature shapes selected by participants; for ‘feature’ composites, participants selected, sized and positioned individual facial features (eyes, noses, mouth, hair, etc.); for the ‘holistic’ EvoFIT, participants repeatedly selected from arrays of complete faces, with ‘breeding’, to ‘evolve’ a face. In all cases, participants worked at their own pace with a suitably-trained police artist or composite operator, with the aim of constructing the best likeness of the face.

Animated sequences of each composite were prepared using caricaturing capabilities within PRO-fit composite software, as per Frowd et al. (2007b). This commercial software exaggerates shape information in composites with respect to an average (reference) face: in this case, an average white male face, since all celebrities used were white males. Initially, 198 coordinates are located on individual facial-features (eyes, brows, nose and mouth), and around the outline of the head and ears, corresponding to those on the average. Exaggerating the differences between coordinates on the composite and coordinates on the average allows a positive (shape) caricature to be produced; de-emphasising the same differences produces an anti-caricature.

Sequences of 21 frames from the 18 celebrity composites were produced as animated GIF files for presentation with Microsoft PowerPoint. Animations were created for the negative range, 0% to -30% (except sketches, for which the range was 0% to -50%); the positive range, 0% to 30% (0% to 50% for sketches); and the full range, which was the same at positive, then at negative. The frame rate was 300 ms, the same as used by Frowd et al., to result in a 6-second duty cycle for full range, and half that for part-range animations. Images looped continuously, with the veridical composite shown first.

For full-range animations, sequences proceeded in the direction of positive caricature up to 30% (50% for sketches), then in the direction of negative caricature, passing through the veridical image at 0% and continuing to -30% (-50%), before increasing the level of caricature again to return to 0%. The positive sequence was the same but omitted the negative part; similarly, the negative sequence progressed from veridical to negative maximum to veridical. An example of the full-range animation may be found at www.uclan.ac.uk/animatedcomposite.

Images were displayed on a computer monitor to dimensions of approximately 8cm wide x 10cm high. Photographs of the target identities used to construct the composites were reproduced on A4 paper, one per page.

Procedure

Participants were tested individually. They were asked to name composites of well-known celebrities, or guess if unsure; it was also mentioned that ‘don’t know’ was an acceptable response. Each person was randomly assigned, with equal sampling, to one type of presentation (static composite / negative range / positive range / full range). Depending on the assigned condition, the relevant set of 18 celebrity composites was presented sequentially, and participants responded with a name (or not) as instructed. Next, the target photographs were presented and participants attempted to name those (static) images, to ascertain familiarity with the target set. The order of identities within each presentation (composites, then targets) was randomised for each person. The task was self-paced, and no feedback was given as to the accuracy of participants’ responses.

Results

The target pictures were scored for accuracy—assigning a ‘correct’ score when an appropriate name was given and ‘incorrect’ otherwise (for no name or mistaken name). Accuracy was high (M > 90% correct in each cell of the design), indicating that participants were very familiar with the celebrities. Composites were similarly scored for accuracy. Correct naming was much lower for composites (M = 30.2%) than for target pictures, but this is the usual situation as composites are error-prone stimuli and are more difficult to recognise than photographs. As Table 1 illustrates, naming accuracy of composites increased from static to negative to positive to full-range presentation.

Table 1 about here

The predictive value of presentation type (static / negative / positive / full-range) on composite accuracy scores was analysed using a hierarchical logistic-regression model. The model involved responses (correct / incorrect) to composites for which participants successfully named the relevant target photographs (946 out of a possible 1008). Used in this way, the analysis takes into account celebrities for which participants were unfamiliar (in effect, by treating unknown target identities as missing data).

The regression model is summarised in Table 1. Presentation type was a significant predictor of naming accuracy. Repeated contrasts found that static and negative conditions were equivalent (p = .19), as were negative and positive (p = .27), but full-range was superior to positive [B = 0.4, $SE(B)$ = 0.2, p = .040, Odds Ratio $Exp(B)$ = 1.5]. Two-tailed Fisher Exact tests, with Bonferroni correction applied for two comparisons (α = .025), found that positive (p = .017, Odds Ratio o = 1.6) and full-range (p < .001, o = 2.4) were superior to static. These data appear to be additive in nature, an observation supported by a significant linear trend from static to negative to positive to full range [B = 0.6, $SE(B)$ = 0.1, p < .001, $Exp(B)$ = 1.9]: higher-order trends were not reliable (X^2 < 1). There were no significant differences for inaccurate responses by presentation[1].

[1]There are two types of inaccurate response for a composite: no name offered and a mistaken name. The latter provides an indication of participants' willingness to provide a name, or guess (response bias), and lower mistaken names, *per se*, indicate stimuli which trigger a more accurate response. Here, inaccurate names were somewhat frequent overall (M = 33.2%), a typical result observed when naming composites (e.g. Frowd et al., 2007b, 2010, 2012). A logistic regression was also run including incorrect responses for those composites where identity was known. Presentation type was not a reliable predictor [$X^2(3)$ < 1].

We next carried out a partial-correlation analysis to provide further evidence for which caricature range was responsible for the naming advantage. For this by-items analysis, correct responses for static naming were subtracted from the relevant response in the other three conditions, to give a difference measure (or *gain*) for each type of animation. The correlation between full-range and positive caricature (r = .88) reduced slightly with a negative-caricature partial (r = .69, p = .002), but full-range and negative (r = .75) decreased substantially with a positive partial (r = -.09, p = .72). This indicates that positive caricature was mainly responsible for the gain observed with full-range animation, a reliable effect as these two partial correlations differ significantly[2] [r(positive) = -.09, r(negative) = .69, r(positive, negative) = .88, Z(15) = 6.8, p < .001].

Discussion

Relative to viewing a static composite, full-range animation greatly increased participants' ability to correctly name the face, replicating the previous work by Frowd et al. (2007b); the increase in effect size from baseline (Constant model) was large (2.4 / 0.5 = 5.2). Frowd et al. proposed that the positive caricature range might be effective by exaggerating facial distinctiveness, information that is potentially valuable for naming, or alternatively that the negative range might be effective by reducing the appearance of errors in composites. The current study provides evidence that the positive caricature range leads to a better probe to memory for composites than does the negative range, but there is evidence that using both negative and positive ranges are beneficial when seen in the same sequence—that is, there was a reliable increase from positive to full-range caricature. The medium-to-large correlation between conditions suggests that animation

---

Similar analyses were conducted on incorrect responses elsewhere in the paper and indicated that mistaken names were not influenced by fixed variables. Frowd et al. (2007b) reached the same conclusion.

[2] For statistical procedure, see Meng, Rosenthal and Rubin (1992).

has a cumulative effect by presentation condition, as does the significant linear trend, replicating the strong relationship between static and full-animation conditions (*r* = .53) reported by Frowd et al. (2007b). The current analysis with separate partial correlations underscores the relatively larger role played by positive caricaturing.

Before interpreting these data any further, we attempt a replication (plus extension) in the following experiment using composites that were constructed using procedures that more-closely model the eyewitness scenario.

Experiment 2: Caricaturing sub-ranges for individual and morphed composites

Method

Design

Faces on which real-life composites are based tend to be less attractive than celebrities' faces: targets chosen for the current study were UK international-level footballers, whose faces were somewhat more rugged in appearance and so were more representative of real-world composite targets. While Experiment 1 used composites generally constructed from familiar faces, this is not usually the case in eyewitness composite construction. Therefore, we took the approach of recruiting composite constructors who were unfamiliar with the targets—in this case, non-football fans—to construct the composites, and football fans to name them.

This study also involved 'morphed' composites. Bruce et al. (2002) demonstrated that combining two or four individual feature composites of the same identity into an image average produces a picture (a morphed composite) that is more identifiable than the average of the individual composites; with averaging, consistent information tends to be maintained, and inconsistent parts (error) diminished. Frowd et al. (2007b) found that estimates of best likeness were positive (*M* = +7%) for morphed composites and negative

(*M* = -10%) for ‘feature’ composites, arguably reflecting the less error-prone nature of morphed images. The naming benefit of morphed over individual composites has been found to extend to composites from a holistic system (Valentine, et al., 2010).

In the present study, we sought to test the prediction that the advantage conferred by caricature animation would again be present for individual composites but would extend to morphed composites. In addition, it was predicted that since morphed composites are inherently less error-prone than individual composites, animating them should produce even better correct naming than animating individual composites.

The experiment involved three groups of participants. Participants in the first group each created a single composite of an unfamiliar footballer, resulting in four composites of each of 10 identities. A second group attempted to name the resulting 40 composites, allowing selection of an intermediate-quality composite for each identity. Selected individual composites and morphs (created from the four composites of the same identity) were then presented to a third group of participants for naming in a between-subjects design for both presentation type (static / negative / positive / full range) and image type (original composite / morph).

Participants

Nine male and 31 female participants were paid £5 to construct a single composite. Their ages ranged from 18 to 54 years (*M* = 25.3 years, *SD* = 8.3 years). One female and seven male football fans, ranging in age from 19 to 38 years (*M* = 25.1 years, *SD* = 5.6 years), volunteered to name the resulting 40 composites. Participants recruited for the main naming exercise were 26 female and 30 male volunteer football fans with an age range of 19 to 62 years (*M* = 33.9 years, *SD* = 12.0 years).

Materials

Targets were 10 front-facing colour photographs of current UK international-level footballers, printed approximately 8cm wide x 10 cm high on A4 paper. Participants who constructed a composite were shown a picture of a randomly-selected footballer; if they reported the face to be familiar, they were shown another randomly-selected photograph until they received one that they did not recognise. Participants then viewed the picture for 60 seconds before freely describing the appearance of the face to the experimenter, who then worked with them and the PRO-fit feature system to construct a likeness of the face. See Frowd et al. (2010) for details of the face-construction procedure. The task was self-paced and each composite took around one hour to construct. This procedure resulted in four composites of each target identity, 40 composites in total.

These 40 individual composites were given to eight football fans to name (with each person seeing a different random order of presentation of composites). To obtain intermediate-quality composites for the main-naming exercise (see Procedure), we chose the item closest to the mean correct naming score for the four composites of each target. PRO-fit software was then used to create negative, positive and full-range animations for these choices, to give four levels of presentation (static / negative / positive / full range). Next, the four individual composites for each target were morphed together using PRO-fit, to produce 10 morphed composites, and this software was used again to create the same three animations, also to give four levels of presentation.

Procedure

The same naming procedure as Experiment 1 was used, except that participants were told that composites were of international-level footballers who play in the UK. Each person was shown 10 intermediate-quality composites or 10 morphed composites, and these were presented in one of the four presentation types. After attempting to name the composites, participants attempted to name the 10 target photographs.

Results

The target photographs were correctly named very well overall (M = 94.5%, SD = 9.5%). Summed correct naming responses for composites are presented in Table 2, and indicate an increase from static composite through to full-range animation, as in Experiment 1; there was also an advantage for morphed over individual composites.

Table 2 about here

Individual responses (correct / incorrect) to composites were again analysed using hierarchical binary logistic regression, see Table 2. The fixed factors were image type (morphed / individual composite) (Step 1), presentation type (static / negative / positive / full-range) (Step 2) and the interaction (Step 3). Image type was a significant predictor of naming accuracy, with morphed composites enjoying more successful naming than individual composites [p = .001, *Exp(B)* = 1.8], but presentation type and the interaction were not reliable. For presentation type, *a priori* testing carried out in the same way as Experiment 1 (including at the same α levels) indicated one significant difference: full-range animations were named significantly higher than static composites (p = .019, o = 1.4). There was also a significant linear trend for this factor [B = 0.4, *SE(B)* = 0.2, p = .016, *Exp(B)* = 1.5], but not higher-order polynomial trends ($X^2 < 1$), again indicating progressively more naming benefit from static to negative to positive to full-range animation, as was observed in Experiment 1.

For a correlational analysis, difference scores were again calculated by-items between static and animation conditions. For individual composites, the correlation between full-range and positive (*r* = .29) was strengthened with a negative partial (*r* = .52, *p* = .15), but the correlation between full and negative (*r* = -.02) was substantially weakened with a positive partial (*r* = -.45, *p* = .23); these partial correlations are reliably different to each other [*r*(positive, negative) = .81, *Z*(7) = 3.2, *p* = .001], replicating the previous result indicating a greater role for positive caricature. The same overall finding emerged for morphed composites: the correlation between full and positive (*r* = .69) was slightly stronger with a negative partial (*r* = .80, *p* = .009) but the correlation between full and negative (*r* = .28) greatly decreased with a positive partial (*r* = -.61, *p* = .08); and these partial correlations differ significantly [*r*(positive, negative) = .79, *Z*(7) = 3.7, $p < .001$]. We note that the above difference between positive-negative partial correlations is much greater for morphs (Δr = 1.41) than for composites (Δr = 0.97), suggesting that the positive-caricature benefit is qualitatively stronger for the former image type.

Discussion

In this experiment, we manipulated (a) facial caricature, the extent to which a face was exaggerated or de-emphasised with respect to the average, and (b) the type of image seen, individual composites or morphed (averaged) composites. We have extended the results of Experiment 1 to composites created entirely of unfamiliar faces, demonstrating again that full-range animation promotes better naming than exposure to static-veridical images. In addition, we have shown that animation, along with improving recognition of individual composites, also promotes better naming of morphed composites—and that, again, morphed composites are significantly more likely to be correctly named than individual composites. Partial correlations once again indicate that the positive range confers most of the animated-caricature advantage; the benefit of caricaturing was stronger for morphed than for individual composites.

So, data from Experiments 1 and 2 support Frowd et al's first explanation: it is the positive range that is most valuable for enhancing composite naming, with recognition being facilitated by increasing distinctiveness of the face. These data argue against Frowd et al.'s (2007b) second explanation: that negative caricaturing works by reducing the appearance of error in composites. If that were the case, then composites should benefit the best from the negative range, but this is not the finding. It is apparent that the negative range is involved to some extent, including the benefit seen from positive to full-range; such an effect may emerge as, sometimes, a composite itself is a positive caricature and so is enhanced by the negative. These findings are considered in more detail in the General Discussion.

In the following study, we attempt to further understand the nature of multi-frame caricature by varying the order in which frames are presented for recognition.

Experiment 3: Varying the presentation order of caricature sequences

The data presented so far suggest that the positive caricature region is mainly responsible for the animation recognition advantage for composites. Another possibility, though, is that this benefit emerges when our face-recognition system extracts the 'direction' in which the transformation occurs—through experiencing the image set in a sequential frame-by-frame order. In the conceptualisation shown in Figure 3, this direction equates to the vector along which the face is caricatured—the dotted line from O to A to z. If such a mechanism were to exist, a vector could be used to recognise the face by calculating the angle between probe image and memory, z .. O .. y: the smaller the angle, the more likely recognition would be to occur. The approach has been used in pattern-recognition research (e.g. Yambor, Draper & Beveridge, 2002); there is also evidence that face discrimination is better in response to changes in angle relative to equivalent changes based solely on distance (e.g. Ross, Hancock & Lewis, 2010).

Figure 3 about here

The idea that processing of an angle might be involved in the caricaturing advantage can be tested by investigating the impact on recognition of randomising the order of frame presentation: a randomised frame-order should disrupt a sequential frame-by-frame extraction of a vector (e.g. O .. A .. z). However, animating frames in a random order produces stimuli that would be too disjointed for participants to view, and so the 'photospread' format was used, by presenting each 21-frame sequence for a composite on a single sheet of paper.

Method

Participants

Participants were 17 female and 13 male volunteers from a local social club in the Lancashire area, UK; ages ranged from 18 to 53 years (*M* = 32.3 years, *SD* = 10.6 years).

Materials

The celebrity set from Experiment 1 was used. The 21 frames containing each composite were printed on single sheets of paper, in sequence, from maximum negative through to maximum positive caricature. This was identical to Frowd et al. (2007b), with three rows of seven images in landscape orientation, except for convenience here printing was done on A4 paper (they used A3). Individual images were approximately 4 cm wide x 6 cm high and the array almost completely filled the page. Out-of-sequence composites were the same, except that frames were printed in different, randomised orders.

Design

Participants inspected composites from one of three presentations (static composite / in-sequence photospreads / out-of-sequence photospreads) in a between-subjects design. To support a vector-direction account, correct naming was expected to be lower for out-of-sequence than in-sequence photospreads; evidence against such an account would see no significant difference in naming between these two presentations.

Procedure

The naming procedure from Experiment 1 was again used, with participants tested individually in a self-paced task. Each person was randomly assigned, with equal sampling, to inspect one of three types of presentation format: individual composites, or photospreads that presented frames simultaneously in order of increasing caricature ('in-sequence') or in a randomised order ('jumbled'). Composites were presented sequentially for naming, as normal, followed by the target photographs. Stimuli were shown in a different random order for each person.

Results

Correct naming of target pictures was very high (*M* = 85.4%, *SD* = 19.8%). See Table 3 for summary of responses to composites and resulting logistic-regression model. Presentation was a reliable predictor of naming accuracy, and repeated contrasts found that in-sequence [*B* = 0.5, *SE(B)* = 0.2, *p* = .024, Exp(B) = 1.7] photospreads were correctly named significantly more often than static images, replicating the multi-frame caricature advantage, but the critical comparison of out-of-sequence (jumbled) versus in-sequence did not differ significantly [*B* = 0.2, *SE(B)* = 0.2, *p* = .29], therefore not supporting a vector-based account for caricature. Out-of-sequence photospreads were named reliably more often than static presentation using a two-tailed Fisher Exact test ($p < .001$, $o = 2.1$).

Table 3 about here

Discussion

Participants more-accurately named both photospreads than static composites, again indicating the benefit of multi-frame caricature as well as replicating Frowd et al. (2007b) using this presentation format. Out-of-sequence and in-sequence photospreads did not differ reliably—although, a linear trend analysis (in the order shown in the table) favoured out-of-sequence over in-sequence [$B = 0.5$, $SE(B) = 0.2$, $p < .001$, $Exp(B) = 1.7$]; the quadratic trend was not reliable ($X^2 < 1$). Overall, the result provides evidence against a vector-direction explanation for the multi-frame caricature effect; support for that would have required the reverse benefit: in-sequence over out-of-sequence.

Having established that a vector-based account of multi-frame caricature is unlikely, we now focus further on an explanation based around positive caricature.

Experiment 4: Reduced-speed multi-frame presentation

In this study, the aim was to estimate the mean exaggeration (frame) which led to recognition when the animation progressed from negative to positive, or the reverse, from positive to negative. We supposed that mean caricature for recognition should be positive and, based on the results of Experiment 3, direction of change should not be important. To more easily locate the frame of recognition, the speed of animation was reduced and participants were asked to stop the sequence when they could name the face.

This method also allowed us to explore whether sequence repetition would help—as frame order seems to be largely unimportant, it should not help as presented stimuli would remain the same with each repeat.

Method

Materials

To make the procedure easier for participants, a subset of 10 celebrity composites from Experiment 1 was used. This subset excluded targets that had previously elicited poor naming in the static and full range, and also those that were repeated across different composite construction methods (and so could result in unwanted inter-item cuing). A bespoke computer program was written to present the composite sequences, as follows.

Animated stimuli could be presented in a 'downward' direction from +50% to -50%, or 'upward' in the opposite direction, from -50% to +50%. To measure the impact of composite sequences alone, each sequence (for a given composite) was made to pause first on the starting frame, only moving on to the next frame if participants could not name it; if they provided a name, the next sequence (for the next composite) was shown. In cases where participants could not name the starting image, single frames were shown for 1.5 s each, and the whole 20-frame sequence would last for 30 s if seen in its entirety.

To explore if repeated exposure to the animation cycle would be advantageous, for which we predicted it would not, two additional conditions were included. Both of these conditions played sequences at the original 300 ms per frame, but differed in their starting frame. The 'normal' animation, which was the same as full-range caricature, started at the veridical image (0% caricature), while the 'max-normal' started from maximum positive caricature (+50%) in half of the trials and from maximum negative caricature (-50%) in the remainder. A summary of presentation type is shown in Table 4.

Ten printed photographs of the targets were used, as in Experiment 1.

Participants

Twenty-three female and 17 male volunteers were recruited from staff and students at UCLan. Their ages ranged from 19 to 61 years (M = 32.3 years, SD = 11.6 years). There were an equal number of participants in each of the four types of presentation. Participants were different to those appearing in the other experiments.

Design

Participants were presented with composites in one of four presentation types (normal / downward / max-normal / upward) and so the design was between-subjects.

Procedure

Participants were tested individually, and randomly assigned with equal sampling to one of the four presentation conditions. They were informed that they would view composites of celebrities' faces: first as static images and then as animated sequences. For each identity, participants were instructed to try to name the static image; if this was not possible, the face would be made to change, and they should press the spacebar to stop the animation as soon as they recognised the face. The starting frame of the first sequence was then presented according to the assigned condition, and participants were asked to try to name it. This frame was at -50% caricature for upward, 0% for normal, +50% for downward, and either -50% or +50% (randomly chosen) for max-normal. If participants could not name the face, the sequence was started: when participants recognised the face, they were instructed to stop the sequence themselves by pressing the spacebar and say the name. If the sequence was not stopped in the normal and max-normal conditions, it ran for five cycles—the same duration as the other conditions (30 s).

For upward, normal, and the portion of the max-normal trials starting at max negative, the animation progressed towards positive caricature; for downward and for the remaining trials for max-normal, the animation progressed towards negative caricature. In the upward and downward conditions, frames were shown for 1.5 seconds, while for normal and max-normal, it was 300ms (to accommodate a constant 30 second total in all conditions). When participants were able to name the face, they were also asked to provide a percentage score reflecting confidence in their own accuracy; the program then moved on to the next sequence without providing any feedback for accuracy of response.

This procedure was repeated for all 10 caricature animations, presented in a different, randomised order for each person. Afterwards, the 10 target photographs were presented in different random orders and participants were asked to name those images.

Results

Participants correctly named the target pictures very well overall (*M* = 93.9%, *SD* = 9.1%). The composites were named correctly by at least one person, either on the basis of static images presented at the start or subsequently during animation. The incidence of correct naming of starting (static) images (first row of data, Table 4) was very low for the maximum anti-caricature (upward), but was much higher for the other conditions. The benefit of viewing animated sequences (difference between first and second rows) was large and fairly consistent across conditions.

Overall effects. A logistic regression model was built for (correct / incorrect) responses (Table 4) with predictors as motion (Step 1), presentation (Step 2) and motion x presentation (Step 3). Motion was a reliable predictor for naming accuracy, and thus seeing the sequence move once again improved naming performance; type of presentation was also reliable, and one simple contrast was significant relative to normal: a naming deficit for upward [*B* = -0.9, *SE(B)* = 0.3, *p* = .001, *Exp(B)* = 0.4].

Table 4 about here

Naming of starting frames. Consider naming of initial (static) images for downward, max-normal and upward (first row of data in Table 4, right-most three columns). It was predicted that upward (-50% caricature) would be worst, downward (+50%) best, and max-normal (-50% for half the time and +50% for the rest) in between. Re-running the logistic model for these static naming responses again emerged a significant predictor for presentation type and two repeated contrasts confirmed the above prediction (*p*s < .01).

Direction of motion. Partial correlations were used to assess whether direction of animation was an important factor for successful recognition. This by-item analysis involved the *gain* in correct-naming with animation for normal, downward and upward conditions. The correlation coefficient was non-significant between normal and upward with a downward partial (*r* = .60, *p* = .09), and also between normal and downward with an upward partial (*r* = .57, *p* = .11). As these coefficients do not differ significantly [*r*(upward) = .60, *r*(downward) = .57, *r*(downward, upward) = .26, *Z*(7) = -0.4, *p* = .70] , this suggests that direction of animation is not involved in recognition with animation.

Confidence ratings. Participants provided a confidence rating after having named a composite. For upward and downward, we calculated mean confidence for identities that were named correctly; there were some missing cells (as not all identities were correctly named by at least one person), but we were able to calculate ratings for the same seven identities by condition. For these data, confidence at having provided a correct name did not differ significantly [*t*(6) = 0.5, *p* = .63] between downward (*M* = 74.2%, *SD* = 25.8%) and upward (*M* = 69.5%, *SD* = 11.5) conditions. The mean level of confidence (for six identities) was 41.2% (*SD* = 8.8%) when participants were incorrect, and higher at 65.5% (*SD* = 13.7%) when correct, a reliable effect [*t*(5) = 2.5, *p* = .044, *d* = 2.1]. Confidence ratings were not significantly different [*t*(5) = 0.8, *p* = .47] when participants correctly named static images (*M* = 68.3%, *SD* = 32.9%) than when they viewed moving caricatures (*M* = 76.4%, *SD* = 11.6%). These all appear to be sensible findings.

Frame of recognition. The next analysis considered correct responses elicited in the downward and upward conditions—bottom row, second and fourth columns of data, Table 4. In both of these conditions, eight of the 10 composites were correctly named, seven of which involved the same target identities in both conditions. For these seven, the mean caricature (frame) at which the sequence was stopped was +7.6% for downward (*SD* = 16.6%) and +16.1% (*SD* = 34.8%) for upward, a non-significant difference [$t(6) = 0.5$, $p = .60$]. The overall mean frame for recognition was positive, at 11.8% caricature (*SD* = 17.8%).

Sequence repetition. In the normal presentation when the animated sequence began, participants always pressed the spacebar before the end of the cycle (12 times in total)—and, they were always correct. This result provides evidence to support the prediction that recognition occurs without the need for sequence repetition.

We now consider max-normal sequences, presentations that proceeded at standard speed (300ms frame rate) in an upward direction for half of the time and a downward direction for the other half. All ten target identities were correctly named at least once during the animated part of the sequence, and the mean correct naming for upward and downward did not differ significantly [$t(9) = 0.6$, $p = .54$]. These results further suggest that direction of sequence presentation is not important. Unlike the normal presentation, however, participants usually stopped the sequence during the second cycle when correctly naming the face. Repetition was probably necessary here since the starting frame (maximum negative or maximum positive caricature) was quite different from veridical, inhibiting recognition to some extent (i.e. making recognition take longer to occur, necessitating presentation of further frames).

Discussion

In this experiment, there was a reliable effect of caricature on static images presented at the start of each trial. When images were maximally anti-caricatured, correct naming was reliably lower than normal; it was highest for maximum positive caricature—slightly but not significantly higher than the veridical image. These data fit part of the pattern found by Frowd et al. (2007b): worst naming for a single, fixed level of negative caricature, and little change for a single, fixed level of positive caricature. However, the earlier study found that a small amount of anti-caricature (10%) promoted slightly more correct naming than the veridical image (0%), a result that was reliable in a sensitive 'cued' naming task. In the upward condition here, the degree of anti-caricature was much greater (-50%) and there was a naming deficit. Thus, while there may be some benefit for representations shifted slightly towards the face used for averaging, more extreme transformations in this direction remove distinctive and important information, reducing performance considerably. Max-normal, in which the two caricature extremes were each presented for half of the trials, appropriately elicited intermediate naming for starting images. When participants viewed animations, irrespective of how likely those images were named statically, correct naming increased by the same reliable amount.

The results also indicate that the direction in which the animation is seen is not important, either negative- or positive-going. This finding clearly supports Experiment 3: order of sequence presentation (in-sequence or random) is not important for a multi-frame caricature benefit. In the current experiment for the normal procedure, recognition occurred without sequence repetition. These data continue to favour the explanation that a caricaturing advantage for composites emerges mainly from one frame in the positive range; it is also independent of direction of travel and does not require frame repetition.

General Discussion

Facial composites produced from traditional ‘feature’ methods tend to be error-prone and so poorly recognised. In this paper, we investigated one technique to facilitate naming: previously, Frowd et al. (2007b) found that observing composites with multiple levels of positive and negative caricature substantially improved correct naming relative to seeing the static composite. We presented four experiments that explored why this technique is an improvement over the normal method of observing a single image.

In Experiment 1, the effectiveness of different sub-ranges of the caricature transform was explored. The positive range was more valuable to naming than the negative range, although there was some contribution from the negative—that is, when viewed in addition to the positive region (in the full-range animation). In this experiment, using composites of celebrities, and the replication in Experiment 2, using composites of footballers, partial correlations between static and full-range animations support the superiority of the positive over the negative range for naming. In Experiment 2, recognition of morphed composites was also enhanced under multi-frame caricature; there was also a replication of the full-range animation advantage (relative to veridical composites) and greater benefit was found overall for the positive over negative range.

Experiment 3 presented multi-frame caricature in a ‘photospread’ format, with the 21 constituent frames printed on a single sheet of paper in either a sequential or a random order. Both of these formats were named reliably better than veridical composites, as found previously, but the same as each other (although the data favoured random orders), providing further evidence that a single frame in the sequence facilitates recognition.

In Experiment 4, successful recognition was found to occur at around the same positively-caricatured frame (M = +11.8%) regardless of whether participants were initially presented with fully-negative or fully-positive caricature, again favouring the relevance of the positive region. Sequence repetition was not important: recognition of the normal sequence occurred before all frames were seen.

Face-space model for caricaturing composites

The robust nature of multi-frame caricature can be explained within a norm- or an exemplar-based model of face space, and so a generic account is given that applies to both. Face-representations are conceptualised in two dimensions, see Figure 1, although the approach applies to multi-dimensional space. Dark circle A is a composite projected into this space, with B and E caricatured versions and C an anti-caricature; grey circles are familiar-face memories. All identities are modelled as single points, but these are likely to represent a region in face space (e.g. Burton et al, 2011). We also argue that face space is different for each person, for which there is good evidence (Burton et al., 2011). Simply, multi-frame caricaturing can be modelled by presenting different representations of a composite along the dashed-line shown; one frame in the sequence (B) will be closest to correct identity (D) and so will be most likely to trigger recognition.

There is an intriguing consequence of this theoretical account: we have a tendency to create composites that are negative caricatures. This is because composite location A is an anti-caricature: it is located closer to the reference face (used for caricature) than is memory D. As such, it benefits from positive caricature. We have argued that this is the general case, but acknowledge that sometimes a composite may be either neutral (B), or a large-positive caricature (E) with recognition facilitated by negative caricature.

At face value, this proposal is at odds with Frowd et al.'s (2007b) finding that we prefer negative caricature as best likenesses for composites (e.g. M = -10% for feature types), as this suggests that composites are the opposite: positive caricature. More recent evidence suggests that we prefer a small degree of anti-caricature for familiar (veridical) faces in general (Allen, Brady & Tredoux, 2009; Hancock & Little, 2011). Also, that the magnitude of these estimates is similar to those obtained from feature composites—Allen et al., for example, found mean estimates in the region of -10 to -15% caricature for photographs of personally-familiar faces. So, anti-caricature estimates are related to face perception more than as a mechanism to reduce the appearance of error in composites.

This reasoning is consistent with results from morphed composites, images that are the average of a number of individual composites and have inherently lower error variance. Frowd et al. (2007b) found that participants prefer slightly positive caricature versions (M = +7%) as best likenesses for morphed composites; being of better-quality than individual images (e.g. Bruce et al., 2002; Valentine et al., 2010; Experiment 2, here), participants attribute less-negative caricature to morphs. Also, being more average, errors are reduced with morphed relative to individual composites and so we model the representation of morphs as being further away from the reference face: they are *less* anti-caricatured. Indeed, results of Experiment 2 suggest that positive caricature is more effective for morphed than for individual composites, which fits with the general idea that the negative range is less important for this type of average representation.

So why might people tend to produce anti-caricatured composites? This is undoubtedly a side-effect of the difficulty in accessing detailed information about facial memory, an issue which psychologists have been aware of for some time (e.g. Davies, Shepherd & Ellis, 1978). The idea is associated with a general forgetting of facial information (e.g. what the eyes looked like, how close were the eyes and brows), an effect which becomes more pronounced as the delay increases from seeing the face (Ellis, Shepherd & Davies, 1980). As a result, constructors select individual features and/or specify inter-relations that are *less* distinctive than the target's. A face constructor may select eyes, for example, that are not quite as distinctive as those seen in a target face. Indeed, after fairly-long intervals, correct naming from feature systems tends to be very low (e.g. Frowd et al., 2005a, 2007a, 2010) and composites themselves have a ratherbland appearance (Frowd et al., 2005a, 2007b). In contrast, systems based more on face recognition than recall, in particular EvoFIT, are now much better at accessing memory. This is evidenced by good correct-naming levels overall (*M* = 46%) for EvoFITs produced from a one day-old memory of an unfamiliar face (Frowd et al., 2012).

Summary and practical implications

What are the practical implications of the research? It has been shown that presenting the full animation sequence, from negative to positive, elicits best naming results. Also, that combining composites into a morphed image, and then animating that image, yields even better correct naming. Full-range animated individual composites and animated morphed composites would therefore appear to be very useful for the police; for example, for publication on wanted persons' web pages, online newspapers and on TV. The technique is permitted for use by UK police (as part of guidelines from the national working party on facial identification) and is referred to in terms of police evidence as an 'enhancement' technique for maximising the effectiveness of eyewitness evidence.

Since publication of the original work, over 100 police officers have been trained to produce animations for PRO-fit and EvoFIT systems; and, the technique has been incorporated into the EFIT-V system (Gibson, Solomon, Maylin & Clark, 2009). Often, the static image is displayed first and, when clicked by the mouse, the animated sequence is initiated. However, observing the face in this way is not quite the same as seeing it animate from the start. The data collected in Experiment 4 (normal animation condition) provide evidence that this police method is likely to be appropriate, allowing people to try to name the original composite before seeing it move. Therefore, full-range normal-speed animated caricature appears to be the best format, and so police should continue to use it.

References

Allen, H., Brady, N., & Tredoux, C. (2009). Perception of ‘best likeness’ to highly familiar faces of self and friend. *Perception*, 38, 1821 – 1830. doi:10.1068/p6424

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

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

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., 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.

Burton, A.M., Jenkins, R. & Schweinberger, S.R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, 102, 943–958.

Calder, A.J., Young, A.W., Benson, P.J., & Perrett, D.I. (1996). Self priming from distinctive and caricatured faces. *British Journal of Psychology*, 87, 141-162.

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, 1, 119-124.

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

Frowd, C.D., Bruce, V., Ross, D., McIntyre, A., & Hancock, P.J.B. (2007b). An application of caricature: how to improve the recognition of facial composites. *Visual Cognition,* 15, 1-31. DOI: 10.1080/13506280601058951.

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

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

Frowd, C.D., 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*. DOI: 10.1002/acp.1662.

Frowd, C.D., Skelton F., Atherton, C., Pitchford, M., Hepton, G., Holden, L., McIntyre, A., & Hancock, P.J.B. (2012). Recovering faces from memory: the distracting influence of external facial features. *Journal of Experimental Psychology: Applied*, DOI: 10.1037/a0027393.

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

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

Hancock, P.J.B., & Little, A.C. (2011) Adaptation may cause some of the face caricature effect. *Perception*, **40**, 317 – 322

Johnston, R.A., Milne, A.B., Williams, C., & Hosie, J. (1997). Do distinctive faces come from outer space? An investigation of the status of Multidimensional Face-Space. *Visual Cognition*, 4, 59-67.

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

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

Lee, K.L., & Perrett, D.I. (1997). Presentation-time measures of the effect of manipulations in colour space on discrimination of famous faces. *Perception*. 26, 733-752.

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. (2004). Face-Space-R: Towards a unified account of face recognition. *Visual Cognition*, 11, 29-69.

Meng, X.I., Rosenthal, R., & Rubin, D.B. (1992). Comparing Correlated Correlation Coefficients. *Psychological Bulletin*, 111, 172-175.

Rhodes, G. (1996). *Superportraits: caricatures and recognition*. Psychology Press: Hove, UK.

Rhodes, G., Carey, S., Byatt, G., & Proffitt, F. (1998). Coding spatial variations in faces and simple shapes: a test of two models. *Vision Research*, 38, 2307–2321.

Rhodes, G. & Leopold, D. A. (2011). Adaptive norm-based coding of face identity. In A. W. Calder, G. Rhodes, M. H. Johnston, & J. V. Haxby (Eds.). *Oxford Handbook of Face Perception*. Oxford: Oxford University Press.

Rhodes, G., & Tremewan, T. (1994). Understanding face recognition: Caricature effects, inversion and the homogeneity problem. *Visual Cognition*, 1, 275-311.

Ross, D., Deroche, M., & Palmeri, T. (2011). Face adaptation: Comparing norm-based and exemplar-based models. *Journal of Vision,* 11, doi: 10.1167/11.11.586.

Ross, D.A., Hancock, P.J.B., & Lewis, M.B. (2010). Changing faces: direction is important. *Visual Cognition*. 18, 67-81.

Tanaka, J.W. & Simon, V.B. (1996). Caricature recognition in a neural network. *Visual Cognition*, 3, 305-324.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology*, 43A, 161-204.

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.

Yambor, B. D., Draper, B.A., & Beveridge R. (2002). Analyzing PCA-based Face Recognition Algorithms: Eigenvector Selection and Distance Measures. In H. Christensen & J. Phillips (Eds.) *Empirical Evaluation Methods in Computer Vision*. World Scientific Press: Singapore.

List of Figures and Tables

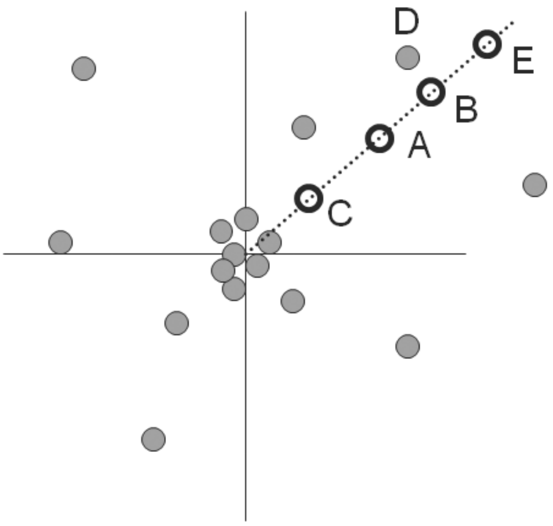

Figure 1. Norm-based face space in two dimensions. Grey circles represent memories (familiar identities) and dark circle A is a composite image, a “probe” to these memories. Exaggerating the features of A away from the centre (norm), along the dotted line, produces positive caricatures, B and E, which in this case B is closer to the correct identity, D. Representation C is a negative (anti-) caricature of A, and thus it has been moved closer towards the norm; C is further away from D and so is harder to recognise than A. Removing the axes and coding items using absolute coordinates illustrates the alternative, exemplar-based face space proposed by Valentine (1991).

Figure 2. Different levels of caricature applied to a composite of former UK Prime Minister, Tony Blair. From left to right, exaggerations are at -30%, -15%, 0% (veridical), +15% and +30% caricature. Negative percentages represent anti-caricatures, zero (0%) the veridical composite, and positive percentages are positive caricatures.

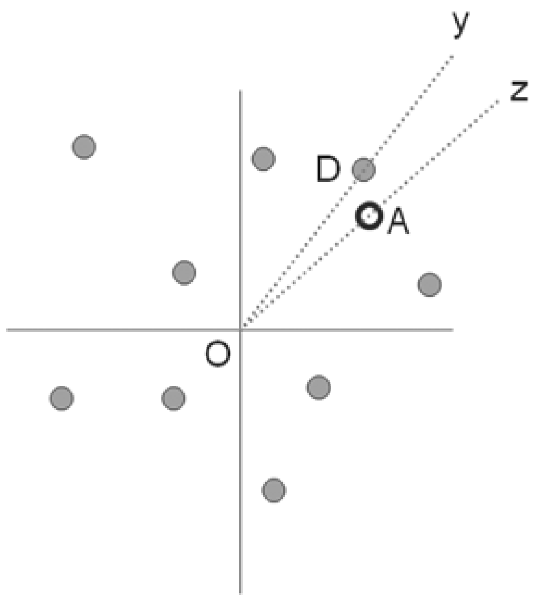

Figure 3. Face space with probe (composite) image A and an existing memory D. In this model, recognition occurs when the angle between A and D (y .. O .. z) is minimal.

Table 1. Advantage of the positive caricature range on composite naming (correct and incorrect) accuracy

Static composite	Negative caricature	Positive caricature	Full-range animation
23.4 (54 / 231)	28.6 (69 / 241)	33.3* (78 / 234)	42.5* (102 / 240)

Note. Figures are percentage-correct accuracy calculated from responses in parentheses: summed correct responses (numerator) and total (correct and incorrect) responses (denominator). These data are for composites for which participants correctly named the relevant target (N = 946 out of 1008). The regression model was significant, $X^2(3) = 21.4$, $p < .001$, $R^2 = .02$ (Cox & Snell) and .03 (Nagelkerke). Hosmer-Lemshow (model fit) $X^2 < 1$. Presentation type [$X^2(3) = 21.1$] and the Constant [$B = -0.8$, $SE(B) = 0.1$, $X^2(1) = 118.7$, $Exp(B) = 0.5$] were significant (ps < .001). See text for more details. *Categories significantly different to each other and from the static composite ($p < .05$).

Table 2. Linear trend by presentation (in the order shown), and the advantage of morphed relative to individual composites

Presentation type				*Image type*	
Static image	**Negative caricature**	**Positive caricature**	**Full-range animation**	**Individual composite**	**Morphed composite**
35.9† (47 / 131)	40.3† (52 / 129)	44.1† (60 / 136)	50.4†* (67 / 133)	35.7 (94 / 263)	49.6** (132 / 266)

Note. Figures are percentage-correct accuracy calculated from responses in parentheses (see Table 1, Note). Data are from individual and morphed composites that participants correctly named the relevant target (N = 529 out of 560). The final model was significant at Step 2, $X^2(4) = 16.5$, $p = .001$, $R^2 = .03$ (Cox & Snell) and .04 (Nagelkerke). Hosmer-Lemshow $X^2(6) = 1.9$, $p = .93$, Dispersion $\phi = 0.3$. Image [$B = 0.6$, $SE(B) = 0.2$, $X^2(1) = 10.2$] and the Constant [$B = -0.3$, $SE(B) = 0.1$, $X^2(1) = 11.7$, $Exp(B) = 0.7$] were significant (ps = .001) predictors. Both Presentation [$X^2(3) = 5.9$, $p = .11$] and Image x Presentation (p = .60) were not reliable. †Significant positive linear trend across the categories in the order shown ($p < .05$). *Category significantly higher than static composite ($p < .05$). **Category significantly higher than individual composite ($p = .001$).

Table 3. The null effect of frame order on correct naming for A4 ‘photospreads’ of caricatured composites

Static Composite	**In-sequence photospread**	**Jumbled photospread**
18.0 (37 / 205)	27.5* (56 / 204)	32.1* (71 / 221)

Note. Figures are percentage-correct accuracy (see Table 1, Note); data are for composites that participants knew the relevant target (N = 630 out of 702). Model, $X^2(2) = 11.6$, $p = .003$, $R^2 = .02$ (Cox & Snell) and $R^2 = .03$ (Nagelkerke). Hosmer-Lemshow $X^2 < 1$. Presentation [$X^2(1) = 11.0$] and the Constant [$B = -1.1$, $SE(B) = 0.1$, $X^2(1) = 113.4$ $Exp(B) = 0.3$] were significant (ps < .005). *Category significantly higher than static composite ($p < .05$).

Table 4. The impact of direction and speed of motion: the disadvantage of upward presentation

	Presentation type			
	Normal	**Downward**	**Max-normal**	**Upward**
Starting frame	veridical composite	max positive caricature	max positive (A), or max negative caricature (B)	max negative caricature
Direction of motion	increasing positive then negative caricature	increasing negative caricature	increasing negative then positive caricature (A), or the opposite (B)	increasing positive caricature
Frame rate	normal (300 ms)	slow (1.5 s)	normal (300 ms)	slow (1.5 s)
Motion type				
Static naming	27.5 (25 / 91)	30.1 (28 / 93)	17.5 (17 / 97)	7.4** (7 / 95)
With motion	40.7 (37 / 91)	51.6 (48 / 93)	40.2 (39 / 97)	29.5** (28 / 95)

Note. Figures are percentage-correct accuracy (see Note, Table 1); data are for composites that participants knew the relevant target (*N* = 752 out of 800). 'With motion' scores are summed responses from both static and moving conditions. The final model was significant at Step 2, $X^2(4) = 61.7$, p < .001, $R^2 = .08$ (Cox & Snell) and $R^2 = .11$ (Nagelkerke). Hosmer-Lemshow $X^2(6) = 4.1$, $p = .66$, $\phi = 0.7$. Motion [$B = 1.0$, $SE(B) = 0.2$, $X^2(1) = 35.3$, $Exp(B) = 2.7$], Presentation [$X^2(3) = 24.1$] and the Constant [$B = -1.2$, $SE(B) = 0.2$, $X^2(1) = 40.7$, $Exp(B) = 0.3$] were significant predictors (*p*s < .001); Motion x Presentation (*p* = .26) was not reliable. **Category significantly lower than Normal (*p* = .001).