An application of caricature: how to improve the recognition of facial composites

Charlie Frowd[1*]
Vicki Bruce[3]
David Ross[2]
Alex McIntyre[2]
Peter J.B. Hancock[2]

[1]Department of Psychology, University of Central Lancashire, UK.
[2]Department of Psychology, University of Stirling, UK.
[3]College of Humanities and Social Science, University of Edinburgh, UK.

[*]Corresponding author: Charlie Frowd, Department of Psychology, University of Central Lancashire, Preston PR1 2HE, UK (email: cfrowd@uclan.ac.uk, Tel: 01772-893439).

Full Article

Running title: Caricaturing facial composites

Abstract
Facial caricatures exaggerate the distinctive features of a face and may elevate the recognition of a familiar face. We investigate whether the recognition of facial composites, or pictures of criminal faces, could be similarly enhanced. In this study, participants first estimated the degree of caricature necessary to make composites most identifiable. Contrary to expectation, an anti-caricature was found to be best, presumably as this tended to reduce the appearance of errors. In support of this explanation, more positive caricature estimates were assigned to morphed composites: representations which tend to contain less overall error. In addition, anti-caricaturing reduced identification for morphed composites but enhanced identification for individual composites. While such improvements were too small to be of value to law enforcement, a sizeable naming benefit was observed when presenting a range of caricature states, which appeared to capitalise on individual differences in the internal representation of familiar faces.

Acknowledgements
The work was funded by a grant from the Engineering and Physical Sciences Research Council, EP/C522893/1. Partial funding was also received from the EPSRC as part of a public understanding of science grant, Future Face, GR/R86034/01. The authors would like to thank the London Science Museum, for allowing us to collect data, the journal reviewers, who made many insightful comments and suggestions, as well as Yvonne Plenderleith, Anna Law and Steve Fields for their assistance with data collection.

Introduction

Facial caricatures are traditionally produced by artists grossly exaggerating a few characteristic features of a face. These cartoon impressions are often quite impoverished and yet remain highly identifiable. While there is a high degree of agreement about which features artists caricature, there are individual differences in the degree of exaggeration applied (Goldman & Hagen, 1978). More consistent approaches to caricaturing are obtained using image processing software (e.g. Benson & Perrett, 1991; Brennan, 1985). In general, the automated approaches exaggerate the shape of a face relative to an average face. The consequence is that those features, plus the relationships between the features, that are distinctive are exaggerated.

Research has found that faces exaggerated with a small degree of caricature using these software programs are generally perceived as better likenesses of a person than the veridical images (e.g. Benson & Perrett, 1991, 1994; Ellis, 1990; Lee, Byatt, & Rhodes, 2000; Rhodes, Brennan, & Carey, 1987). For example, Rhodes et al. (1987) used a line-based caricature generator devised by Brennan (1985) to create both caricatures and anti-caricatures of familiar faces (staff and students) at 25% intervals from 0 to +/-75%. They found a general preference for drawings seen with a positive caricatures and, by interpolation, estimated the level of best likeness to be 16%. They also demonstrate that a 50% caricature led to faster reaction times than veridical images, which in turn was faster than a 50% anti-caricature. Benson and Perrett (1994) used line drawings of famous faces but found that the mean estimate of best likeness was higher than Rhodes et al., interpolated at 42%. They also found that participants were faster to respond to drawings that had been caricatured at optimal levels obtained from their initial perceptual task.

Benson and Perrett also found that identification accuracy increased when caricaturing line drawings, a result that has been replicated (e.g. Rhodes & Tremewan, 1994; Tanaka & Simon, 1996), and also with slightly more realistic-looking line-drawings (Mauro and Kubovy, 1992). Caricatures may therefore serve as a ‘superportrait’, with elevated identification relative to veridical images. In general, caricatures of line drawings tend to be identified as well as veridical drawings (and sometimes a little better) but anti-caricatures tend to be less identifiable (e.g. Rhodes et al., 1998; Rhodes & Tremewan, 1994).

Photographs of faces are also preferred when caricatured, though the degree of distortion required to produce the best likeness is lower. Benson and Perrett (1991) report a 4% level (by interpolation) for a set of famous faces; Ellis (1990) reports 6%. Unlike line drawings, veridical photographs are highly identifiable, and this may make it difficult to promote a caricature advantage. However, by briefly presenting stimuli to reduce overall performance, Lee and Perrett (1997), and later Lee et al. (2000), report an identification advantage for a 50% caricature presented for 67ms (and conversely, an identification decrement for a 50% anti-caricature). A caricature advantage for photographs of faces has also been found by Benson and Perrett (1991) in a face-to-name matching task.

The general benefit of a faster response and/or better naming for caricatures of faces, and the reverse effect for anti-caricatures of faces, can be explained within Valentine’s (1991) face recognition framework in two ways. The first holds that faces are coded by their distance and direction from a central prototype, or norm. Caricaturing increases the distance without affecting the direction, making the face more distinctive (Stevenage, 1995). The alternative does not require a norm and posits that faces are held as values on a set of dimensions. Moving a face away from the average has the effect of increasing its distance from other faces, making it more identifiable (Lewis & Johnston, 1998, 1999). Neither model appears to accommodate all the available data (Byatt & Rhodes, 1998). However, the finding that caricatures work better for reduced representations of faces, such as line drawings, than for photographs has been modelled by Lewis (2004) in terms of a lower number of effective dimensions of variability. The current work considers facial composites of the type used in police investigations, which generally lie somewhere between photographs and line drawings in their quality and might therefore be expected to show an intermediate degree of caricature effect.

The composites in question are those produced by witnesses to, or victims of, serious crime. These images are usually constructed by selecting individual facial features from a kit of parts – eyes, nose, mouth, etc. – thus building up a 'composite' image[1]. While these 'feature' systems are typically computerised, facial composites are also produced with the assistance of a sketch artist, who draws out the face by hand; there are also other non-feature systems that are also emerging, as discussed below. Irrespective of the system, however, modern composites tend to be identified infrequently (e.g. Brace et al., 2000; Bruce et al., 2002; Davies, van der Willik, & Morrison, 2000; Frowd et al., 2005a, 2005b, in press, under revision; Koehn and Fisher, 1997). It is likely that requiring witnesses to select individual facial features is at variance with the natural way that faces are perceived, as wholes (Davies, Shepherd and Ellis, 1978), and therefore the resultant faces are less than optimal. They tend to appear rather bland (partly due to the need to blend bits of different photographs together) and insufficiently distinct to allow others to identify the person represented. It is therefore of interest to test the effects of caricaturing: from an applied view, since any increase in recognition would interest the police, and from a theoretical perspective, to see whether they do indeed produce a maximal effect somewhere between the levels reported for photographs and line drawings.

A further class of image tested here is *morphed composites*. Bruce, Ness, Hancock, Newman & Rarity (2002) showed that such 'morphs'[2] of four individual attempts at a composite produced recognition at least as good as the best individual composite. It is believed that this works by averaging out individual inaccuracies in each composite, while emphasising any consistencies. However, such averaging inevitably produces an image that looks even more bland than the source composites (see Figure 3 for an example). Our prediction, therefore, is that morphed composites should show a stronger caricature effect than the originals.

In Experiment 1, participants adjusted the degree of caricature to make each presented image most identifiable to another person. In Experiment 2, we tested whether these empirical caricature settings did promote better recognition. In Experiment 3, improved recognition was found when participants were presented with multiple levels of caricature for each composite. A replication is reported in Experiment 4, using a different set of composites.

INTRODUCTION TO EXPERIMENTS

The following experiments featured facial composites produced from previous research projects (e.g. Bruce et al., 2002; Fields, 2005; Frowd et al., 2004, 2005b, under revision) using four different composite production systems, which differ in image quality and therefore in the potential effects of caricaturing.

For brevity we omit full details of the procedures used to construct the composites, which are available in Frowd et al. (2005b). In brief, participants looked at a photograph of a target face for 1 minute, described his face, and used one of four methods to construct a composite from memory with the assistance of an experienced computer technician (operator) or sketch artist. Each composite had been constructed using procedures that broadly matched those used in police work and included a Cognitive Interview, to assist recall of the face (Geiselman et al., 1987); open ended construction sessions, to promote the best visual likeness; and artistic enhancement, to improve the likeness of a composite (e.g. Frowd et al., 2005b; Gibling & Bennett, 1994).

Three different techniques were used to construct the composites. The first was a highly experienced UK police sketch artist. This person drew out the composites by hand using pencils based on feature shapes selected by witnesses. The technique involved initially working on the proportions of the face and then progressively increasing the level of detail. The second was E-FIT and PRO-fit, typical computerised 'feature' systems in current UK police use. Composites were produced by witnesses selecting individual facial features from an internal kit of facial parts. As they are very similar to each other, and produce very similar composites (Frowd et al., 2005a, 2005b), they were treated equivalently here and collectively referred to as *EP-FIT*. The third system was EvoFIT, a recognition-based system recently released to the UK police (Frowd et al., 2004). With EvoFIT, composites are 'evolved' through a process of selection and breeding of whole faces. EvoFIT images are near photographic in quality, those from EP-FIT somewhat less so, as the component parts are sourced from different photographs, while sketches tend to have more limited shading information (see Figure 2 for examples). On Lewis's (2004) account of caricature effects, we might therefore expect EvoFITs to show least caricaturing and sketches the most.

With the exception of the final experiment, which sought to replicate results found elsewhere in this paper and whose composites were constructed as realistically as possible in the laboratory, the other composites varied in the degree of familiarity that participant-witnesses had with target faces (familiar / unfamiliar); the famousness of the target (well-known celebrity / personally familiar); and the length of time participant-witnesses waited between inspecting a target face and constructing a composite (no delay / 3-4 hours). The main criterion for selecting the composites was that the target faces from which they were constructed should be highly identifiable to participants here. Details of the particular targets and systems used may be found in the relevant experiments below.

EXPERIMENT 1: PERCEPTUAL DATA

Experiment 1A and 1B sought the level of caricature necessary to maximise the recognition of individual composites and morphed composites, respectively. This followed the basic design of Benson and Perrett (1994), using a computer program which allowed participants to adjust the degree of caricature for each composite via a graphical slider. Participants were asked to locate the setting which made each composite as identifiable as possible.

EXPERIMENT 1A: COMPOSITES

Method

This part explored the level of caricature necessary to maximise the recognition of 18 famous face composites, six taken from each of three different production methods. The methods were sketch, EP-FIT and EvoFIT, as described above and illustrated in

Figure 2. Note that while it is possible to caricature the intensity or *texture* information in a face, which has been shown to help automatic recognition systems (e.g. Craw, Costen, Kato, Robertson & Akamatsu, 1995), the procedure used throughout this paper caricatured only the shape information.

The average face used for caricaturing was from EvoFIT (e.g. Frowd, Bruce, Plenderleith & Hancock, 2006), being a blend of 72 male faces with an average age of 30 years, using 252 coordinate points to match key landmarks, as shown in Figure 1. The average age is a good match for the composites used here.

Figure 1. An average of 72 faces with coordinate landmarks superimposed. The coordinates were used as a reference points for caricaturing.

Equivalent landmarks were identified for each of the 18 composites, using PRO-morph[3], which also produced the caricatures, by moving the points away from the average, or towards it for the anti-caricatures. Participants were allowed to adjust the degree of caricature from -50% (anti-caricature) to +50% (caricature) in steps of 5% using a computer mouse.

It was expected that participants would assign a moderate level of positive caricature to the composites, somewhere between the 4.4% reported by Benson and Perrett (1991) for photographs and the 16% reported by Rhodes et al. (1987) for line drawings. Given that the composite sketches had limited shading, resembling line drawings, they should have the highest caricature setting, while EvoFITs, most like photographs, should have the lowest.

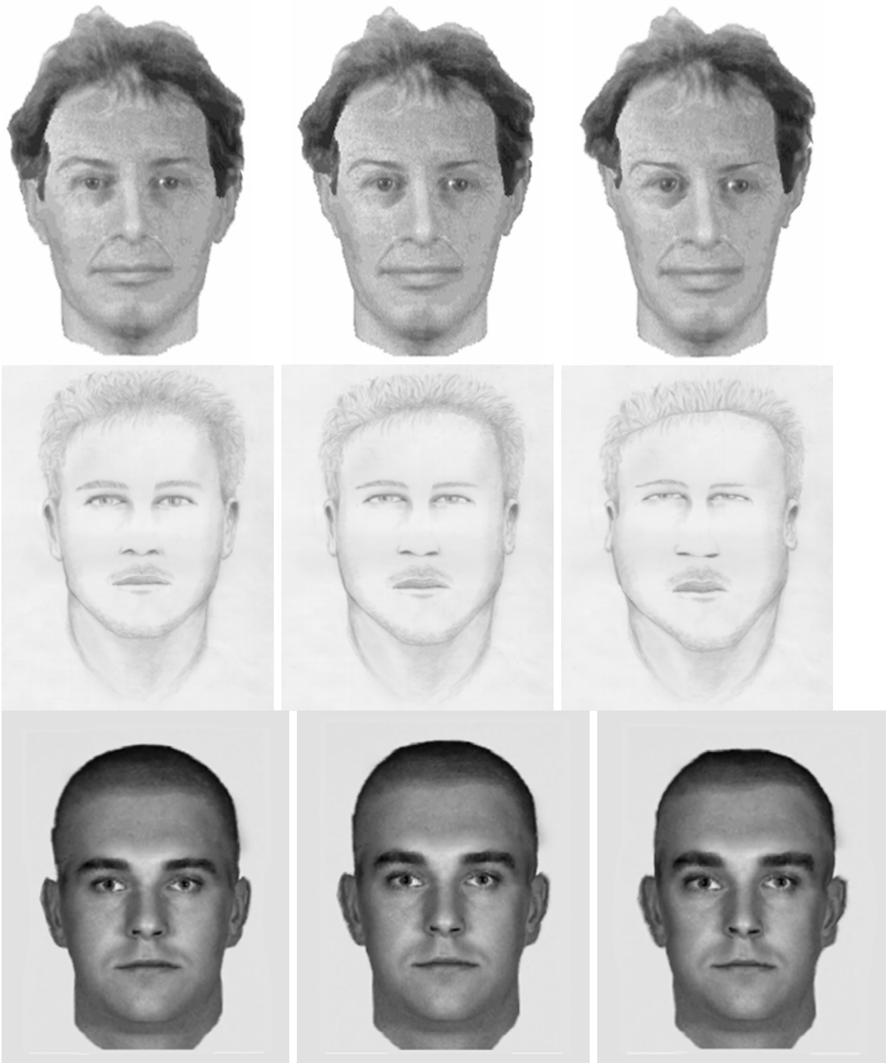

Figure 2. Example composites (centre) and caricatures (left image -50% caricature, right +50% caricature). Included were an EP-FIT of Tony Blair (top), a sketch of Brad Pitt (middle) and an EvoFIT of Robbie Williams (bottom).

Materials

The composites were of 18 famous faces, comprising six composites each from the sketch artist, EvoFIT and EP-FIT. Presentation was on a Hi-Grade laptop at approx. 7cm wide x 10cm high.

Participants

Participants were 59 visitors to the London Science Museum, and 33 were female (*M* = 28.2 years, *SD* = 12.8). Of these, 22 visitors inspected EP-FITs, 18 inspected sketches, and 19 inspected EvoFITs. Participation was voluntary.

Procedure

Participants were tested individually and inspected six composites produced from one system (sketch, EvoFIT, or EP-FIT). Each person was told that they would see a series of famous face composites and for each to try to make the face as identifiable as possible, so that another person would best recognise it. The computer software was introduced, which presented the composite, the name of the celebrity, and the graphical slider used to modify the face. After a short demonstration of the program, participants worked sequentially through the composites and set the slider as requested. They were also instructed to indicate any faces that were unfamiliar (this was recorded). Unknown to participants, the sense of the slider was randomised such that half the time a movement to the right gave a positive caricature. When finished, participants were asked whether they recognised the face transform being applied, which very few did, and they were then debriefed about caricaturing and the purpose of the experiment. The order of composite presentation was randomised for each person.

Results

The celebrity composites used were well-known to participants, being familiar on average 93.8% (*SD* = 12.7%) of the time. The following analysis concerns caricature settings only for celebrities that were familiar. The average caricature setting for the 18 composites was -14.7% (*SD* = 16.0%); that is, an anti-caricature. Table 1 shows that 14 of the 18 (78%) composites have negative average values. By system, sketches were attributed the greatest anti-caricature, at -33.1%, followed by EP-FITs at -10.1%, and EvoFITs at -4.0%, a significant effect of system by an ANOVA, $F(2,15) = 8.8$, $p < .01$. A by-subjects analysis indicated that these average settings were reliably different from a zero caricature level for both sketches, $t(17) = 12.6$, $p < .001$, and EP-FITs, $t(21) = 4.4$, $p < .001$, but not for EvoFITs, $t(18) = 1.4$, $p > .1$; similarly, by-items analyses compared with zero were significant for sketches, $t(5) = 10.6$, $p < .001$, marginally significant for EP-FITs, $t(5) = 2.2$, $p < .1$, but were not significant for EvoFITs, $t(5) = 0.6$, $p > .1$.

Table 1. Mean participant slider settings for the famous face composites produced by the three methods of construction. The data reveals a consistent preference for an anti-caricature. Values are caricature settings expressed as a percentage.
* Significant compared with a zero caricature, $t > 4$, $p < .001$, by-subjects.

Celebrity	EP-FIT	Celebrity	Sketch	Celebrity	EvoFIT
Brad Pitt	-9.3	Andre Agassi	-31.8	Anthony McPartlin	3.4
Michael Caine	-26.8	Brad Pitt	-40.6	David Beckham	-15.8
Nicholas Cage	-0.9	David Beckham	-41.1	Michael Owen	-30.3
Noel Gallagher	-11.9	Michael Owen	-31.2	Robbie Williams	16.3
Tony Blair	-15.9	Noel Gallagher	-25.0	Tim Henman	-3.5
Woody Allen	4.3	Robbie Williams	-24.2	Will Young	4.1
M	-10.1*		-33.1*		-4.0
SD	10.8		11.2		12.1

Discussion

The data clearly indicate a preference for negative caricature settings, which was strongest for sketches, then EP-FITs, with EvoFITs showing no reliable difference from veridical. This suggests that the most identifiable representation was perceived as one where the facial features were generally de-emphasised. The ordering is as predicted, but in the opposite direction.

We propose two explanations for why anti-caricatures were preferred. The first is that witnesses may have produced an image that is already caricatured in their attempt to recall the face. The second is that they will, inevitably, make some errors in their construction and that anti-caricaturing may reduce these. Caricaturing would simply emphasise any such inaccuracies, making identification harder, so one prediction is that the worst composites should be given the strongest anti-caricaturing, as observers attempt to reduce the inaccuracies. As discussed above, morphed composites show relatively enhanced recognition, arguably because these inaccuracies are averaged out. If the second explanation is correct - which was anticipated since composite quality tends to be quite low - the preferred caricature settings for such morphed composites should be positive, to help emphasise the aspects of the face agreed upon by the individual composites. If the first explanation is correct, then the morphed composites should still be caricatures, if rather more weakly since different people may caricature different aspects of the face, and the preferred setting should remain negative.

EXPERIMENT 1B: COMPOSITES AND MORPHS

Materials

The morphed composites were produced using the landmarks as in the previous experiment, morphing each to the average shape and then averaging the four images. Bruce et al. (2002) found that the rated likeness and the recognition of composites were improved through morphing.

The PRO-fit composites used here were of six male and three female members of staff in Psychology at the University of Stirling, each constructed by four different witnesses to produce a set of 36 composites (Bruce et al., 2002; Fields, 2005). ABM's PRO-morph software was deployed both to average together the four composites from the same target and to caricature the resulting nine morphed and 36 original composites (from -50% to +50% caricature in steps of 5%). Example stimuli can be seen in Figure 3.

Figure 3. An individual composite (top row) and a morphed composite (bottom row), an average of four composites, of a member of staff in Psychology. Shown are -50% caricature (left), veridical (centre) and +50% caricature (right).

The design and procedure were as for Experiment 1A except that participants were told that they would be inspecting composites of staff in Psychology (rather than of famous faces). It was expected that the composites would again be judged best at -10% overall. In contrast, the morphed composites were expected to be assigned more positive caricature settings.

Participants

Six female and four male staff and students in Psychology at the University of Stirling volunteered (*M* = 38.9 years, *SD* = 11.7).

Results

The mean estimate for the original composites was -11.6% (*SD* = 9.6%), which differed significantly from a zero caricature, both by-subjects, $t(9) = 3.8$, $p < .005$, and by-items, $t(35) = 4.4$, $p < .001$. An average negative value was assigned to 30 of the 36 items (83%); these data do not differ significantly from those reported for the EP-FITs in Experiment 1A, $t(30) = 0.4$, $p > .1$, by-subjects. In contrast, the average setting for the morphed composites was +7.2% (*SD* = 7.7%), a significant increase from veridical by-subjects, $t(9) = 3.0$, $p < .02$, but not by-items, $t(8) = 1.3$, $p > .1$, despite the mean caricature setting being positive for 7 out of the 9 items.

Discussion

The average caricature setting for the morphed composites was positive, though significantly so only in the by-subjects analysis. The original composites showed the same negative preference as the EP-FITs in Experiment 1A, despite using different targets - those personally familiar rather than famous. While not ruling out some positive caricaturing in the original composites (explanation one above), the results were consistent with the second explanation, that the negative caricaturing chosen in

Experiment 1A helped to reduce errors, and that these errors were also reduced by the morphed averaging process. With these errors reduced, a positive caricature is chosen. We now test whether caricaturing can actually improve recognition for these images.

EXPERIMENT 2: NAMING

Bruce et al. (2002) found that morphed composites were on average more identifiable than the individual composites from which they were composed. Thus, for the current work, one would not only expect elevated naming for the morphed composites relative to the individual composites, but also that a negative caricature would make the individual composites more identifiable, and a positive caricature would make the morphed composites more identifiable (and vice versa). These possibilities were explored in the following two experiments.

EXPERIMENT 2A: STATIC NAMING I

In this experiment, we compare naming for the morphed composites, and the individual composites from which they were produced, at three different levels of caricature: -10%, 0 and +10%. These values were chosen as round figures close to the average optimal values found in Experiment 1B for each type. We expected naming to be better at +10% for the morphs and -10% for the originals. The design was between-subjects for degree of caricature but within-subjects for image type.

In Experiment 1B, there were four composites for each of nine target faces, or 36 composites in total. To obtain a more manageable set for naming, only one composite was used from each set of four per target face, chosen as the item which elicited an intermediate level of naming (based on existing naming data from Bruce et al., 2002, and Fields, 2005).

Research has consistently demonstrated that composites are poorly named, typically around 15-20% (e.g. Brace et al., 2000; Davies et al., 2000; Frowd et al., 2005a, 2005b) and consequently analyses tend to be hampered by low values. To overcome this we included a cued naming condition, which required participants to name the composites a second time, after knowing the identities of the target set; it has the benefit of generally elevating naming levels.

Participants

Seven male and 17 female staff and students at the University of Stirling, different from those in Experiment 1B, were paid £1. Their age was 20 to 28 years ($M = 22.0$, $SD = 1.8$).

Materials

Three naming booklets were used, each containing nine individual composites and nine morphed composites at one level of caricature (-10% / 0% / +10%). Each image was printed on a single sheet of A4 paper in monochrome using a high quality printer at approx. 7cm (high) x 5cm (wide), as were nine staff target photographs.

Procedure

Participants were tested individually and asked to name a set of composites of members of staff in Psychology. They were randomly assigned, with equal sampling, to one of the three testing booklets. The nine individual composites and nine morphed composites were presented sequentially and participants were requested to provide a name for each where possible, in their own time. Participants were then asked to name the target faces from the photographs, again presented sequentially. Next, participants were unexpectedly asked to name the composites a second time. For this *cued* naming condition, the composites were presented in the same order as uncued naming. We applied an *a priori* rule to only permit participant data when at least half the target photographs were named, since low target naming would produce low composite naming. Data from two additional participants were discarded on this basis. The order of all items was randomised for each person.

Results

As expected, the spontaneous (uncued) naming rate was higher for the morphed composites (M = 38.4%, SD = 17.3%) than the individual composites (M = 23.1%, SD = 13.1%). The data show the predicted pattern of results for the negative caricature manipulations, as indicated in Figure 4. Notably, there was a naming advantage for anti-caricatures of composites, effect size d = 0.4, and, to a much greater degree, a naming deficit for anti-caricatures of morphed composites, d = 2.0 (relative to veridical composites and veridical morphs respectively). The participant naming scores were subjected to a mixed-factor ANOVA, which showed an effect of image type, $F(1,21) = 16.6$, $p < .001$, indicating the general advantage for morphed composites, and caricature type, $F(2, 21) = 3.8$, $p < .05$. However, these factors interacted, $F(2,21) = 7.7$, $p < .005$, as the anti-caricatured morphed composites were named significantly less often than the other two conditions, $p < .001$ (and also as the morphing advantage did not extend to the -10% caricature condition, $p > .1$) [4,5]. Analysis by-items showed only a significant interaction, $F(2,32) = 3.6$, $p < .05$, as the anti-caricatured morphed composites were named significantly less often than the other two sets of morphed composites, $p < .01$.

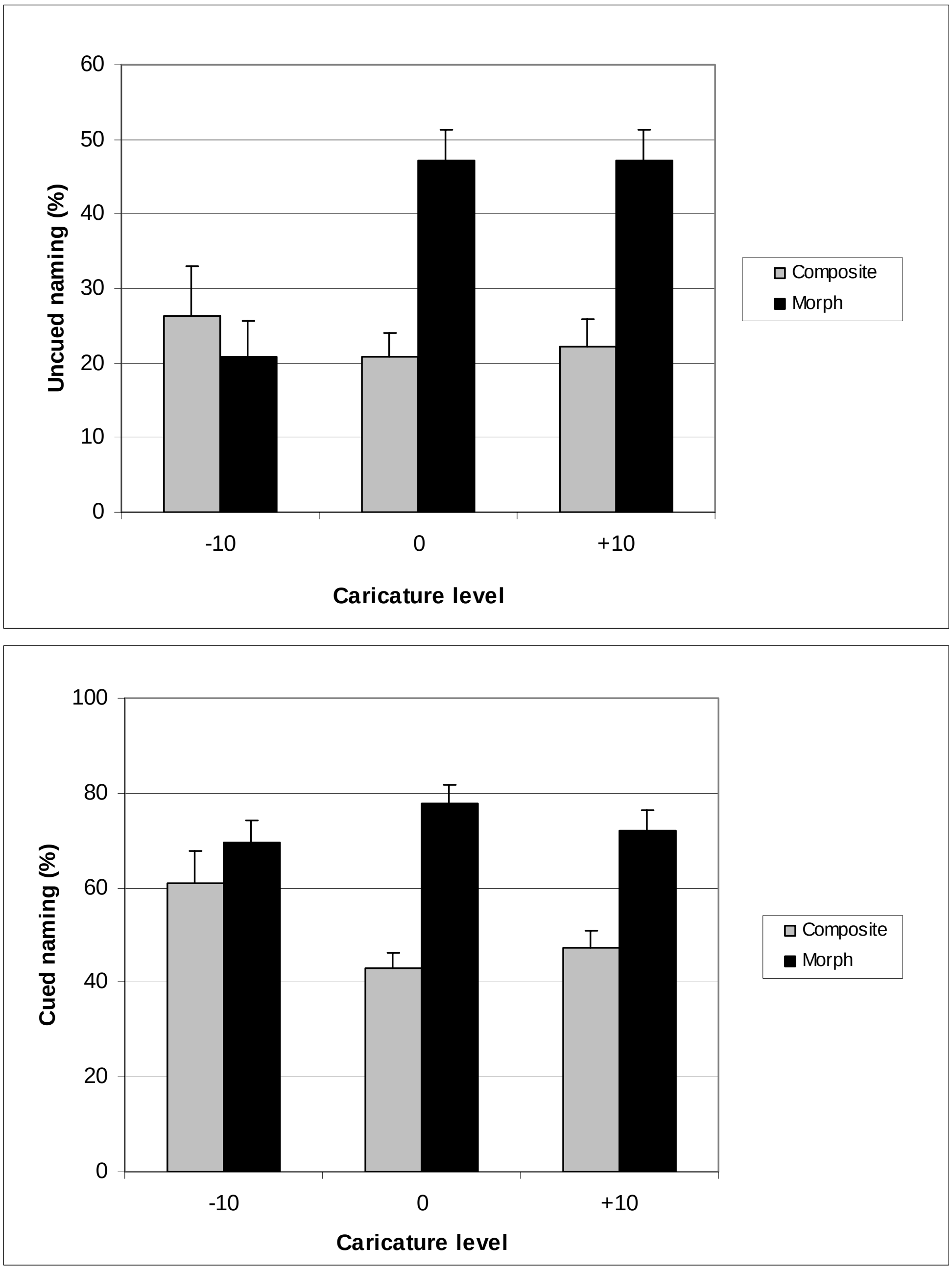

Figure 4. Spontaneous (uncued) naming (top) and cued naming (bottom) for the individual and morphed composites. The data presents a large naming deficit for antic-caricatured morphed composites (uncued) but a large naming benefit for anti-caricatured individual composites (cued). Error bars are standard errors of the means.

Naming was higher when cued (*M* = 61.8%, *SD* = 30.4% vs. *M* = 30.7%, *SD* = 29.3%). As can be seen in Figure 4, performance was also higher for morphed composites (*M* = 73.1%, *SD* = 14.6%) than individual composites (*M* = 50.5%, *SD* = 17.6%). The differences across the morphed composites conditions were minimal, but there was a sizeable increase in naming for anti-caricatures of individual composites, $d = 1.1$. These data were significant for composite type, $F(1,21) = 23.9$, $p < .001$, but not caricature type, $F(2,21) = 0.6$, $p > .1$. However, the interaction approached significance, $F(2,21) = 2.1$, $p < .1$, as anti-caricatured composites were named significantly better than veridical composites, $p < .05$ (also, the morphing advantage did not apply to the anti-caricature condition again, $p > .1$).

An items analysis showed only trends in the same directions, indicating relatively large variability between items. This inter-item variability was investigated further by adding caricature level as a factor to the ANOVA. The composites were divided into low (M = -16.9%) and high (M = 7.7%) caricature sets, based on a median split calculated separately for the original and morphed images. The mean naming rate was 22.9% (SD = 13.1%) for the low caricature set and 36.5% (SD = 16.9%) for the high. An ANOVA showed an effect of caricature level, $F(1,21) = 9.7$, $p = .005$, by-subjects, with a large effect size, $d = 0.9$; there were no significant interactions by caricature level.

Discussion

Our predictions were confirmed in part. Firstly, for the uncued data, anti-caricatures of morphed composites were named significantly worse than veridical. Secondly, for the composites, anti-caricatures were generally better named than veridical composites, though the difference was only significant under cued naming.

Experiments 1B and 2A suggest that a morphed composite is a more accurate representation of a target face than an individual composite, thus supporting the work of Bruce et al. (2002). The experiments also suggest that the identification of morphed composites can be reduced through anti-caricaturing, as with photographs of faces. This explanation is consistent with the positive caricature setting being preferred for morphed composites in Experiment 1B, and conversely that the original composites were better recognised when given their preferred negative setting. Overall there is evidence that poorer quality composites, those with lower naming, have more negative caricature settings. This fits with the explanation that the preference for negative caricatures reflects an attempt to reduce inaccuracies in the composites.

In the next part, we explore whether the increase in naming for negatively caricatured PRO-fit composites generalises to other types of composites: those produced by Sketch and EvoFIT.

EXPERIMENT 2B: STATIC NAMING II

Method and Procedure

Experiment 2A found a naming advantage for PRO-fit composites caricatured at -10% using the cued procedure. In the current experiment, we explored whether this result would generalise to other types of composites, namely the sketches and EvoFITs of Experiment 1A. This time though, the degree of caricature used was the mean chosen by participants for each image (those listed in Table 1) rather than the overall average. This was the method used by Benson and Perrett (1994) to demonstrate a caricature effect with line drawings, and should produce the most identifiable representation for each composite and thus a larger effect size for naming.

One composite from Experiment 1 received a mean rating very close to zero and was therefore omitted. This resulted in a set of 17 composites (six composites each for sketch and EvoFIT, and five for EP-FIT). The design was within-subjects for composite system and between for the two levels of caricature: veridical and ‘optimal’, which is negative for many items . The procedure of Experiment 2A was repeated to name the composites, except that participants were initially told that the composites were of celebrity faces.

Materials
There were two testing books of 17 faces, one with veridical composites, the other with caricatured composites. A set of 12 target photographs were similarly printed (12 because 5 faces were used for two different composite types - refer to Table 1).

Participants
Twenty-seven male and 37 female staff and students at the University of Stirling volunteered. Their age ranged from 17 to 57 years (M = 27.0, SD = 8.8).

Results
For the initial, uncued condition, participants correctly named on average 39.2% (SD = 8.0%) of the veridical and 39.5% (SD = 8.2%) of the (anti-)caricatured composites. Performance between different composite systems was also similar (Table 2) and the ANOVAs showed neither significant effects nor an interaction.

Table 2. Uncued and cued naming for veridical (V) and 'optimally' caricatured (C) composites. The data illustrates small differences with uncued naming, but more sizeable increases for distorted versions of both sketch and EvoFIT composites in the cued condition. Values are percent correct.

	EP-FIT		Sketch		EvoFIT	
Naming	C	V	C	V	C	V
Uncued	46.9	45.6	31.8	33.3	41.1	39.6
Cued	68.8	70.6	64.6	53.6	71.9	62.5

Cueing improved naming, to 68.3% (SD = 13.8%) for optimal caricatures, compared with 62.1% (SD = 11.1%) for the unaltered images, d = 0.43. As Table 2 illustrates, the (anti-)caricature advantage held for both sketch and EvoFITs, but not EP-FITs. An ANOVA showed an effect of system, $F(2,124) = 13.2$, $p < .001$, a marginal effect of caricature type, $F(1,62) = 3.6$, $p < .1$, and no interaction, $F(1,62) = 1.4$, $p > .1$. T-tests provided weak evidence that both sketches and EvoFITs were named better when presented as optimal caricatures, $p < .1$. An items analysis showed an effect only of caricature type, $F(1,14) = 6.0$, $p < .05$, supporting the general naming benefit for the optimal caricatures. In summary, there is evidence that, relative to veridical composites, optimal caricatures of both sketches and EvoFITs promoted more correct names in the cued paradigm. Unlike Experiment 2A, there was no significant relationship between the caricature setting for individual composites and their (uncued) naming rates.

Discussion
In this experiment, we investigated whether a naming benefit could be observed when deploying appropriate levels of caricature to the famous face sketches, EP-FITs and EvoFITs. The caricature setting used this time was not the overall average perceptual estimate, as in Experiment 2A (-10% for all composites), but the average *individual* item estimates, some of which were positive and others negative. Unfortunately, these arguably more 'optimally' caricatured composites were no better named than veridical composites in the uncued task, and there was only weak evidence of a naming benefit

in the cued task. For the latter paradigm, correct naming was found to be higher for enhanced versions of both sketch and EvoFIT composites relative to veridical, but not for the EP-FITs, although these presented a weak anti-caricature advantage in Experiment 2A. In general, these effects are clearly subtle, in spite of a medium effect size, and were only present in the more sensitive, but less forensically-relevant, cued naming task.

EXPERIMENT 3: MULTIPLE CARICATURE STATES

It is apparent from Experiment 1 that there are large differences in the extent by which composites were caricatured to produce the most recognisable representation (see Table 1). Such differences between items are typical in research on facial composites (e.g. Ellis, Shepherd & Davies, 1975; Frowd et al., 2005a), facial caricature (e.g. Benson & Perrett, 1991) and other areas in face perception (e.g. Little & Perrett, 2002; Rhodes, 1994; Shapiro & Penrod, 1986; Suzuki, Hoshino, & Shigemasu, in press; Valentine, 1991; Yasuda, Bedard, Mizokami, Kaping, & Webster, 2005). One possible explanation is that there are individual differences between observers in the way that faces are represented internally, perhaps dependent on the range of faces that are known. Such variation is one explanation for the other-race effect (Valentine, 1991). Another possibility is that individual composites vary inconsistently: some features might be exaggerated, while others are underemphasised or just wrong. In either case, presenting a range of different caricature levels might improve recognition. If different individuals prefer different levels of caricature, everyone should see a representation that is best for them. If different bits of an image look best at different levels of caricature, then presenting a wide range should maximise the opportunity for one of them to trigger recognition. This possibility is explored in the following experiments.

EXPERIMENT 3A: A MOVING CARICATURE

Method

Our initial approach was to present composites with a variety of caricature settings in the form of animated GIF images. To do this, an image sequence was used for each composite that spanned the positive (+30% EP-FITs and EvoFITs, and +50% for sketches) and negative extremes (-30% and -50% respectively) in steps of 5%. Sequences were set to repeat every 6 seconds (i.e. at a frequency of 1/6Hz); note that a higher repetition rate was found to be inappropriate, since faces were perceived as making very odd expressions.

The design was between-subjects for presentation type (veridical / animated). This time, all 18 composites from Experiment 1A were presented for naming, 6 per composite system (EP-FITs / Sketch / EvoFIT). As this experiment, and those that follow, was primarily concerned with effects which were large enough to be of practical value for police work, only the uncued version of the naming task was administered.

Materials

Two sets of stimuli were required, static composites and moving caricatures, and both were presented on a laptop at a size of approximately 8cm (wide) by 12cm (high). A set of 13 target photographs were printed on A4 paper, as before.

Participants

Twenty-two students at the University of Stirling volunteered, half were male, and none had taken part in Experiment 2B. Their age spanned from 19 to 40 years (*M* = 26.0, *SD* = 6.0).

Procedure

Participants were tested individually. They were randomly allocated, with equal sampling, to either the static presentation condition, and saw the 18 veridical composites, or the motion condition, and saw a set of 18 animated caricatures. As before, they were told that the composites were of celebrity faces and to provide a name for each where possible. Microsoft PowerPoint was used to present the items, which were initially given a random order, and presented in either a forward or a reverse sequential order, with an equal number of participants for each (a fully randomised item presentation was carried out for the replication in Experiment 4). The presentation of stimuli was self paced. Afterwards, participants named the target photographs.

Results

The naming rate for the veridical composites was 28.8% (*SD* = 8.2%), compared with 42.4% (*SD* = 17.4%) for the animated versions, a ‘very large’ effect size of 1.0. An ANOVA shows an effect of composite system, $F(2, 40) = 23.4$, $p < .001$, and of animation, $F(1,20) > 50$, $p < .001$, but no interaction, $F(2,40) = 1.51$, $p > .1$. All systems demonstrated an advantage for animation, as illustrated in Figure 5. T-tests showed this to be significant for sketches, $p < .001$, and marginally for EP-FITs, $p < .1$. The items analysis indicated a marginally significant effect of animation, $F(1,15) = 4.2$, $p = .06$; both system and the interaction were non-significant, $F < 2$, $p > .1$. Again the items analysis is weaker, suggesting inter-item variability. A correlation between caricature setting and (static composite) naming rate across all 18 items is significant, $r = 0.53$, $p < 0.05$. Although this is partly driven by differences between the three composite types, it is not entirely so: for example, one EvoFIT was set at -29% caricature and was named only once when static and eight times (36%) when moving.

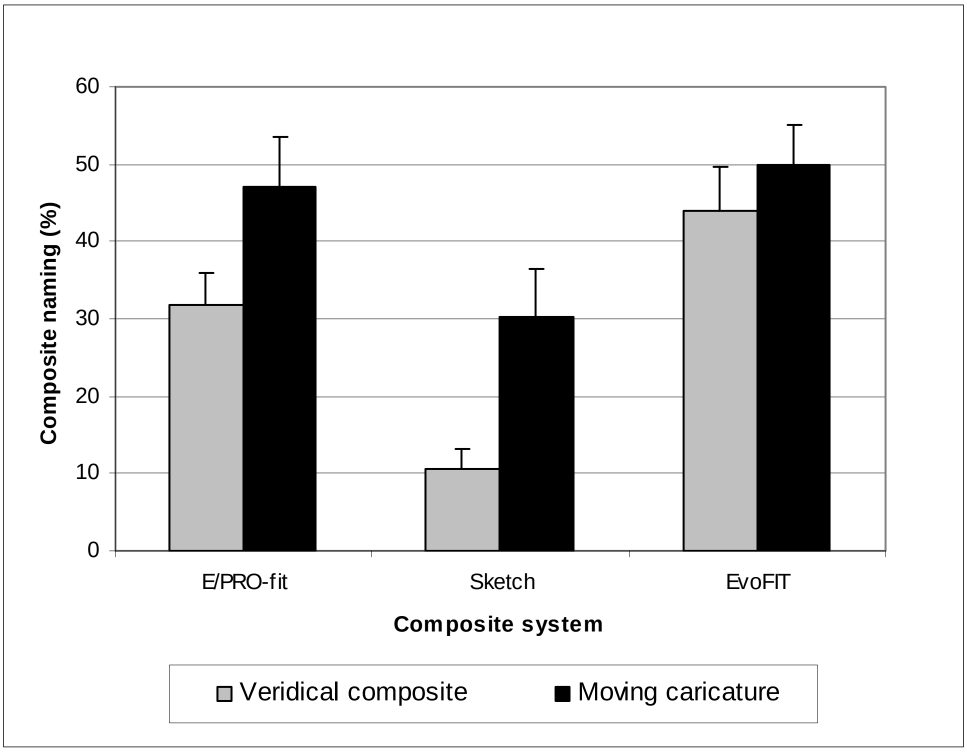

Figure 5. The advantage of a moving caricature. Error bars are standard errors of the means.

Discussion

Naming was higher in the animated caricature condition, especially for sketch, which did worst in the veridical condition[6]. While there are clear differences between the different composite systems as represented here, there is some evidence in support of the hypothesis that negative caricatures are set for poorer quality composites in general: in the relatively weak items analysis and the overall correlation between caricature setting and naming rate.

While the improvement afforded by the animation is consistent with either of the hypotheses, namely individual differences between recognisers or between composites, it is also possible that it is due simply to the motion being more engaging, thus promoting greater attention. This alternative explanation is explored in Experiment 3B.

EXPERIMENT 3B: A MOTION EFFECT?

Method and Procedure

To test the possibility that the naming advantage for the animated caricatures in Experiment 3A was merely a product of motion, as opposed to caricature, the previous experiment was repeated using a photospread-type presentation of static caricature images in place of the animation. The photospreads contained the individual images from the caricature animations for each celebrity. Thus, the identification of single static composites was compared with the identification of photospreads of caricatures.

The design was between-subjects for presentation type (static composite / photospread) and the naming procedure of Experiment 3A was followed, except that materials were presented on paper rather than on a laptop, and a caricature photospread was used instead of the animation. If the original hypothesis was correct, that multiple caricature settings promote identification, then the naming level found here for the photospreads should be the same as that found for the animated caricatures in Experiment 3A; but, if motion itself is of benefit, then the photospreads should be named the same as the static composites.

Materials

The static composites and caricature photospreads were printed at 7x5cm using a good quality monochrome printer. Each static composite was printed in the centre of an A4 sheet; the photospreads were printed on A3 paper (landscape), ordered by increasing positive caricature. A set of 13 target photographs were used from Experiment 3A.

Participants

Twenty-four students at the University of Stirling volunteered, half male. Their age ranged from 18 to 59 years (M = 31.6, SD = 11.6). These participants were different to those in Experiment 2A and 3A.

Results

Naming of static composites was 29.2% (SD = 15.0%), compared to 37.0% (SD = 10.7%) for the caricature photospreads: levels that were very similar to the equivalent conditions in Experiment 3A. Combining data from both experiments, an ANOVA revealed a significant main effect for presentation type (single composite / caricature animation or photospread), $F(1,42) = 7.5$, $p < .01$, but neither experiment (3A / 3B), $F(1,42) = 0.4$, $p > .1$, nor the interaction, $F(1,42) = 0.5$, $p > .1$, approached significance. An analysis by-items was similarly significant for presentation type, $F(1,34) = 8.6$, $p < .01$, but not experiment, $F(1,34) = 0.0$, $p > .1$, nor interaction, $F(1,34) = 0.4$, $p > .1$. These data suggest that caricature rather than animation was responsible for benefits in naming relative to a veridical composite. However, there is a low correlation, $r = -0.05$, between the naming improvement in Experiments 3A and 3B by composite, hinting that there might be two different mechanisms at work.

Discussion

This experiment explored the mechanism underlying the naming benefit for the animated caricature procedure in Experiment 3A. It was found that the presentation of a photospread of static caricatures led to the same significant improvement relative to veridical composites as the animated caricatures in the previous experiment. Therefore, presenting a range of caricature states would appear to be elevating identification, as originally proposed.

EXPERIMENT 4: REPLICATION

The data presented so far provide evidence for enhancing composite identification using a moving caricature sequence. The purpose of Experiment 4 is a replication with a different set of composites, using the PRO-fit system. Specifically, we sought to verify that (a) the overall perceptual caricature estimate would be approximately -10%, as found for the EP-FITs in Experiments 1A and 1B, (b) there would be no significant difference in uncued naming between veridical composites and static caricatures, as in Experiment 2A and 2B, and (c) the moving caricature advantage of Experiment 3A would replicate.

Method
The facial composites were taken from a recent test of a novel feature presentation interface to PRO-fit (Frowd et al., under revision). In this work, 24 witnesses inspected a photograph of one of 12 unfamiliar footballers for 1 minute and each constructed a single composite 2 days later. For the current study, we used composites from the 12 witnesses using the standard interface. See Figure 6 for examples.

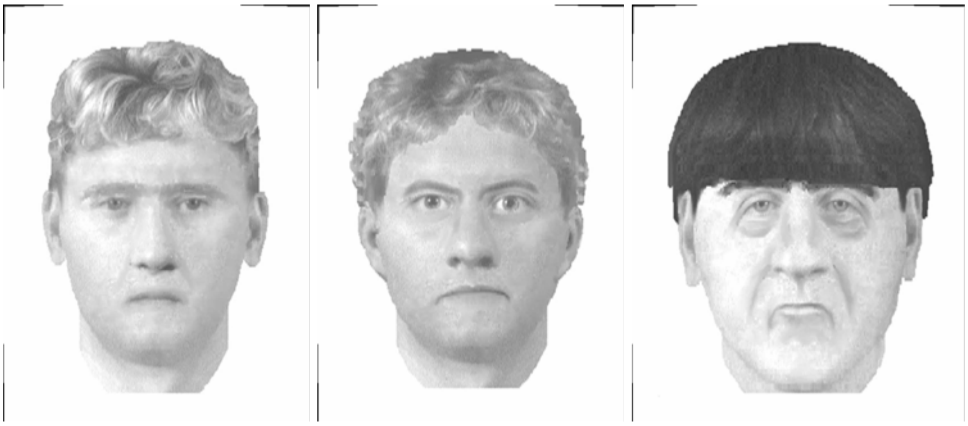

Figure 6. Example footballer composites produced by participant-witnesses. From left to right, they are of Paul Scholes, Ole Gunner Solskjaer, and Peter Beardsley.

The first stage was to collect perceptual ratings of best likeness, as in Experiment 1A. This allowed a three way between-subjects comparison of the original composites, ‘optimal’ (anti-)caricatures and moving caricatures. We expected little difference between the two static conditions but improved recognition for the moving caricatures. We used bespoke software to allow a fully random presentation order for all stimuli.

Participants
Two different groups of football fans volunteered, both sampled from staff and students at the University of Stirling. For the perceptual task, there were 10 males and two females, aged from 22 to 51 years (M = 34.1, SD = 9.4). For the naming task, there were 52 males and 2 females, aged from 18 to 28 years (M = 21.0, SD = 2.2).

Procedure
Part one followed the design and procedure of Experiment 1A to provide average caricature levels necessary to make the composites as identifiable as possible to another person (refer to Table 3) for use in the following naming task. One item was given a mean level of 0% and so appeared identical in the two static conditions, the others were rounded to the nearest 5% as before.

For naming, participants, self-identified as football fans, were told the composites were of generally well-known players competing at an international level. They were randomly assigned in equal numbers to one of the three conditions (veridical composite / static caricature / moving caricature) and asked to name the composites where possible. The program animated the moving composites between -50% and +50% caricature levels in 5% steps with a cycle duration of 6s. Naming was then attempted for the 12 target photographs. The order of presentation of composites and targets was randomised for each person.

Results

The average caricature setting from part one was -11.9% (*SD* = 16.6%), significantly different from zero, $t(11) = 2.5$, $p < .05$, by-subjects and by-items. As Table 3 reveals, the individual settings spanned from -30.0% to +19.2%.

Table 3. Average caricature settings (percent) for the footballer composites.

Footballer	**Caricature**
Tony Adams	6.3
Peter Beardsley	-18.8
Dennis Bergkamp	7.5
Roy Keane	-16.7
Steve McManaman	-30.0
Emmanuel Petit	-21.7
Wayne Rooney	19.2
Paul Scholes	-7.5
Teddy Sherringham	0.0
Alan Smith	-29.5
Ole Gunnar Solskjaer	-29.1
Zinedine Zidane	-22.1
M	-11.9
SD	16.6

Figure 7 shows the results for naming. An ANOVA confirmed an effect of presentation type, $F(2,51) = 5.3$, $p < .01$. Fisher post-hoc tests showed a significant advantage for the moving condition, $p < 0.005$, over both others, with the same ‘very large’ effect size $d = 1.0$ found in Experiment 3A, and no significant difference between the two static conditions. Items analyses also showed a significant effect of presentation type, $F(2,22) = 23.2$, $p < .001$, and of animation, $p < .001$. All items showed a gain in recognition rate in the moving condition. A correlation between the caricature setting and the gain in recognition afforded by movement for each item was significant, $r = -0.66$, $p < .05$, indicating that the composites given the most negative caricature settings gained most by movement. Adding caricature level as a factor to the ANOVA, using a partition about zero, produced a significant effect of caricature on naming, with the low rating group averaging 37.3% (*SD* = 23.6%) and the high group 48.1% (*SD* = 26.1%); the effect size was medium level, $d = 0.4$.

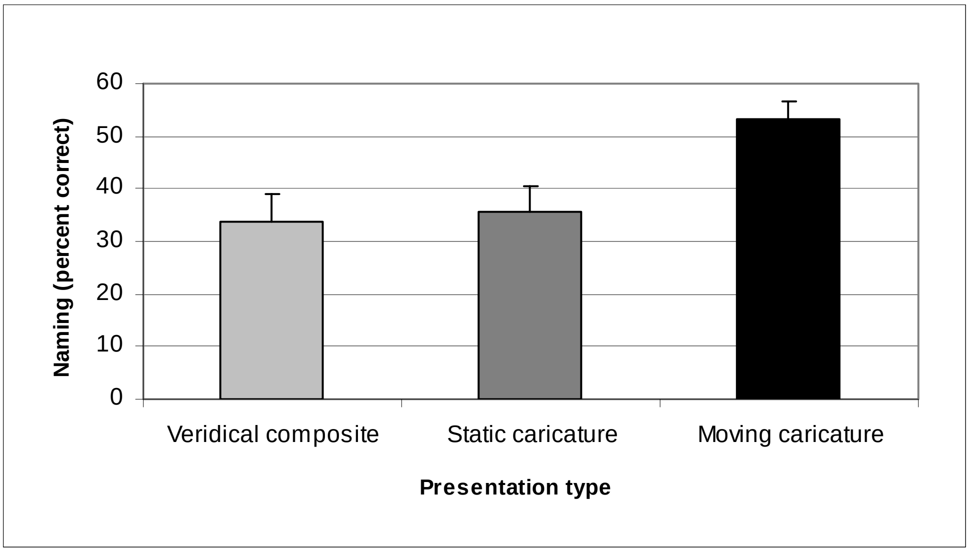

Figure 7. The benefit of a moving caricature for the footballer composites set. Error bars are standard errors of the means.

Discussion

Experiment 4 was designed to replicate some of the basic effects observed in this paper using a different set of facial composites. The PRO-fit composites were constructed of footballers using a realistic design that included a 2 day retention interval and targets that were unfamiliar to participant-witnesses. The preferred caricature settings found in part 1 averaged -12%, very similar to the -10.1% and -11.6% for EP-FITs in Experiments 1A and 1B. As before, however, recognition rates for the static caricatures showed no significant improvement over the veridical images, while there was a large increase for the moving caricatures. All the hypotheses were therefore met.

General Discussion

The normal function of a facial composite system is to render an image from the memory of a witness or victim of crime. Unfortunately, there is good evidence that even using modern construction procedures and systems, composites are poorly identified (Brace et al., 2000; Bruce et al., 2002; Davies et al., 2000; Frowd et al., 2005a, 2005b, in press, under revision; Koehn & Fisher, 1997). Therefore, attempts to make composites more identifiable are both theoretically interesting and of practical benefit to law enforcement. This paper explores the possibility of improving the identification of composites by exaggerating their distinctive features through an automated caricature procedure.

Initially, participants were allowed to set the degree of caricature that would make a set of celebrity composites most identifiable to another person. The results suggested that the best representation was not a caricature, but an anti-caricature, which we interpret as reducing the appearance of errors. The results also suggested that sketches promoted a greater degree of anti-caricature (-33%) than the EP-FITs (-10%) or the EvoFITs (-4%). In the next experiment, the previous perceptual task was repeated using a set of morphed composites, images where inaccuracies were likely to be reduced through averaging. This time, estimates were positive (7%) for the morphed composites. In Experiments 2A and 2B, anti-caricaturing led to a weak recognition benefit for individual composites, but a much stronger recognition deficit for morphed composites. In Experiment 3A, all caricature states were presented via an animated sequence, and this proved to be very beneficial to naming. We then verified that this advantage was a function of caricature rather than motion, and finally replicated the basic static and animation effects.

At the onset of the study, it was hypothesised that caricaturing the shape component of composites might help to make them look more distinctive, more individuated and therefore more identifiable. While caricaturing no doubt improves the distinctiveness of a face (e.g. Lee, Byatt, & Rhodes, 2000), the influence of caricature on shape inaccuracies in composites was unexpected. The composites used in the experiments here were each constructed from a person's memory, and as such are likely to contain some inaccuracies; otherwise, the composites would be highly recognisable, which we know they are not (Bruce et al., 2002; Fields, 2005; Frowd et al., 2004, 2005b). These errors are likely to have emerged from both an imperfect selection of facial features, such that their shapes will be inaccurate, and by an imperfect positioning of the features on the composite face, such that the spatial relationship between the features will be inaccurate. While caricaturing a composite will exaggerate the distinctive features, and potentially make the face appear more identifiable, it will also exaggerate the errors, which will tend to reduce identification. In contrast, an anti-caricature will tend to reduce both incorrect and distinctive aspects of the face. Thus, there appears to be a trade-off between these two influences and reducing the errors in general through anti-caricaturing wins. There is some evidence from several of the experiments that less well identified composites have more negative caricature settings.

Bruce et al. (2002) have argued that averaging a number of composites together produces a more accurate representation, a morphed composite, since the errors that people make when constructing composites are not correlated and so cancel. The data collected here provide further support for this notion. A small overall positive caricature setting was assigned to the morphed composites (Experiment 1B) and, while a positive caricature setting did not significantly increase naming, there was a significant decrease in naming following an anti-caricature (Experiment 2A). This pattern of results is typical of that found for naming of photographs of faces (Rhodes et al., 1998): anti-caricatured faces are worse than veridical, and caricatured versions are at least as good as veridical images. This suggests that morphed composites function more like photographs of faces than do individual composites.

Naming was found to be higher for composites assigned more positive caricature ratings for both personally familiar faces (Experiment 2A) and footballers (Experiment 4), though not for the famous faces in Experiment 2B. There is suggestive evidence also from Experiment 3A that better composites are more positively caricatured. These results are consistent with the finding that morphed composites were associated with more positive caricature ratings and better identification and the hypothesis that the negative caricatures chosen were an attempt to reduce apparent errors in feature shapes and/or their spatial relationships. While the type of caricature carried out here involved both of these components, as is the normal procedure when caricaturing shape information, it is possible to caricature each component separately: by exaggerating either the shapes of the features or the distances between them. Preliminary work in this area revealed broadly similar caricature estimates by shape and by spatial relation, suggesting similar perceived errors in both aspects

In Experiments 3A and 4, a sizeable increase in uncued naming was observed using a moving caricature format which presented a range of caricature states; the same result was observed when the caricature states were presented statically

(Experiment 3B). This approach was originally motivated by the presence of large individual differences in preferred caricature settings in Experiment 1. One explanation is that the memory of a familiar face differs between people and that these memories may be triggered more effectively at different levels of caricature: hence the benefit of presenting a range of caricature states. An alternative is that different parts of a composite face may be constructed more accurately than other parts. For example, in one of the composites, the shapes of the eyes may be constructed well, but not the distance between them. In this case, identification is likely to benefit from presenting both positive and negative caricature levels, since the former would exaggerate the accurate eyes and the latter would reduce the appearance of an incorrect eye placement.

These two possibilities lead to slightly different predictions (Richard Kemp, personal communication). The individual difference explanation would suggest that one particular caricature level would be best for each person; the within-composite variability one that the entire sequence would be best since multiple levels should help. In principle, these two possibilities could be explored by repeating the composite naming exercise with reduced range animations, for example from -50% to -40%, -40% to -30%, etc. The presence of a bimodal naming distribution about a zero caricature would be indicative of the latter, the within-composite variability explanation, but the lack of power in naming studies make firm conclusions unlikely. In general across these experiments, the by-subjects results are stronger than those by-items, suggesting relatively more variation between items than participants. The finding in Experiment 3A that the composites with the most negative caricature settings gained the most from the multiple level presentation also argues for within-composite effects. The highly significant improvements found in Experiments 3 and 4 might suggest that both mechanisms are at work.

Experiments 3A and B found that the improvements in naming from moving caricatures or static arrays of the whole range were very similar. Which would be preferable in practice would depend on context, with a static array required for a poster and movement better on screen. There is a hint from the lack of correlation in improvements between the two presentation modes that there are two different underlying mechanisms, conceivably related to the two explanations of why the effect occurs at all. Further study will be required to confirm and if so explain this.

Lewis (2004) argued that reduced representations of faces, such as line drawings, benefit from stronger caricaturing because there are relatively fewer dimensions of variability. On this basis, we predicted that sketches might get stronger caricaturing than EP-FITs, which would be stronger than the most photographic-like EvoFITs. The results from Experiment 1 confirmed the ordering, but in the opposite direction, with sketches most negative. This could be related to the differences in type of image, but could also relate to the differences in composite accuracy, since this set of sketches is relatively poorly named. An experiment with more directly comparable composites would be needed to resolve this. Again, both explanations might contribute, with the reduced representation of the sketches encouraging a rather larger shift in caricature to remove the inaccuracies.

To conclude, the work indicates that composites constructed from a person’s memory contain errors, which can be reduced in part by anti-caricaturing or by averaging together composites of the same target face. Considerable benefit to naming was observed by presenting a range of caricature states. Questions remain about the underlying theoretical explanation for the effect but our data provides good evidence that the identification of suspects could be increased substantially were composites to be presented using the novel animated format, for example on TV crime programs such as Crimewatch in the UK.

References

ACPO(S) (2000). *National Working Practices in Facial Imaging*. Association of Chief Police Officers (Scotland) Working Group, unpublished document.

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

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.

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

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.

Craw, I., Costen, N., Kato, T., Robertson, G., & Akamatsu, S. (1995). *Automatic face recognition: Combining configuration and texture*. Proceedings of the International Workshop on Automatic Face and Gesture Recognition (pp. 53–58).

Davies, G.M., Ellis, H., & Shepherd, J. (1978). Face identification: The influence of delay upon accuracy of photofit construction. *Journal of Police Science and Administration*, 6, 1, 35-42.

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

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

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.

Ellis, H.D. (1990). Developmental trends in face recognition. *The Psychologist*, 3, 114-119.

Ellis, H.D., & Shepherd, J.W. (1996). Face recall – methods and problems. In S.L. Sporer, R.S. Malpass & G. Koehnken (Eds.). Psychological issues in eyewitness identification (pp. 87-116). Hillsdale, NJ: Lawrence Erlbaum.

Ellis, H.D., Shepherd, J., & Davies, G.M. (1975). Use of photo-fit for recalling faces. *British Journal of Psychology*, 66, 29-37.

Fields, S. (2005). Improving facial composites through a physical reinstatement of context. Unpublished undergraduate dissertation, University of Stirling.

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

Frowd, C.D., Bruce, V., Ness, H., Thomson-Bogner, C., Peterson, J., Mcintyre, A., & Hancock, P.J.B. (under revision). Parallel approaches to composite production. *Ergonomics*.

Frowd, C.D., Bruce, V., Plenderleith, Y., & Hancock, P.J.B. (2006). Improving target identification using pairs of composite faces constructed by the same person. *IEE Conference on Crime and Security* (pp. 386-395). IET: London.

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., 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.

Goldman, M., & Hagen, M. A. (1978). The forms of caricature: Physiognomy and political bias. *Studies in the Anthropology of Visual Communication*, 5, 30-36.

Hagen, M. A., & Perkins, D. (1983). A refutation of the hypothesis of the superfidelity of caricatures relative to photographs. *Perception*, 12, 55-61.

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.

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

Lewis, M.B., & Johnston, R.A. (1998). Understanding caricatures of faces. *Quarterly Journal of Experimental Psychology*, 51A, 321-346.

Lewis, M.B., & Johnston, R.A. (1999). A unified account of the effect of caricaturing faces. *Visual Cognition*, 6, 1-42.

Little, A., & Perrett, D. (2002). Putting beauty back in the eye of the beholder. *The Psychologist*, 15, 28-32.

Mauro, R., & Kubovy, M. (1992). Caricature and face recognition. *Memory & Cognition*, 20, 433-440.

O'Toole, A.J., Bartlett, J.C., & Abdi, H. (2000). A signal detection model applied to the stimulus: Understanding covariances in face recognition experiments in the context of face sampling distributions. *Visual Cognition*, 7, 437–463.

Purtle, R.B. (1973). Peak shift: a review. *Psychological Bulletin*, 80, 408-421.

Rhodes, G. (1994). Secrets of the face. *New Zealand Journal of Psychology*, 23, 3-17.

Rhodes, G., Brennan, S.E., & Carey, S. (1987). Identification and ratings of caricatures: implications for mental representations of faces. *Cognitive Psychology*, 19, 473-494.

Rhodes, G., Byatt, G., Tremewan, T., & Kennedy, A. (1996). Facial distinctiveness and the power of caricatures. *Perception*, 25, 207-223.

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., & Moody, J. (1990) Memory representation of unfamiliar faces. *New Zealand Journal of Psychology*, 19, 70-78.

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

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

Shepherd, J.W., & Ellis, H.D. (1996). Face recall methods and problems. In S.L. Sporer, R.S. Malpass and G. Koehnken (Eds.), *Psychological Issues in Eyewitness Identification* (pp. 87-115). Hillsdale, NJ: Lawrence Erlbaum.

Stevenage, S.V. (1995). Can caricatures really produce distinctiveness effects? *British Journal of Psychology*, 86, 1, 127-146.

Suzuki, A., Hoshino, T., & Shigemasu, K. (in press). Measuring individual differences in sensitivities to basic emotions in faces. *Cognition*.

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.

Yasuda, M., Bedard, K., Mizokami, Y., Kaping, D., & Webster, M. A. (2005). Adaptation and individual differences in categorical judgments of faces. *Journal of Vision*, 5, 832, http://journalofvision.org/5/8/832/, doi:10.1167/5.8.832.

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

Footnotes

[1] There is an unfortunate inconsistency in the literature over the term 'facial composite', which was used by Young, Hellawell & Hay (1987) to describe a novel face produced from the top and bottom halves of two different photographs. In the forensic literature, and that used here, it refers to a face image of a suspect, typically made up of many component parts but also including sketches (e.g. ACPO(S), 2000; Davies et al., 2000; Frowd et al., 2005a, 2005b; McQuiston-Surrett et al., in press; Shepherd & Ellis, 1996); note also that the term 'composite' was in forensic use (e.g. by Davies, Milne & Shepherd, 1983) prior to Young et al. (1987).

[2] 'Morph' is another ambiguous term in the literature, meaning a change in form, but it has also come to mean an average of two or more images, typically faces, since each is morphed to the average shape. We have adopted the term 'morphed composite' (e.g. Bruce et al., 2002).

[3] PRO-morph is a software component of the PRO-fit facial composite system marketed by ABM in the UK. PRO-morph was used here for both caricaturing (all experiments) and averaging the individual composites to produce morphed composites (Experiment 1B).

[4] An analysis of incorrect names generated by participants was considered, since these data may provide an indication of guessing; also, composites with lower incorrect names can limit a waste of police time. However, in all the experiments presented here, the data suggested that caricaturing exerted only a weak influence on incorrect name production. For simplicity, these data are omitted.

[5] An analysis by target naming was conducted to check for the presence of a sampling bias. In spite of randomly assigning participants to testing booklets, chance differences in the naming of the target photographs may still occur, and this itself may lead to differences in the naming of the composites. Overall, target naming was very high, at over 98% correct at each caricature level, and did not differ significantly, $F(2,24) = 0.04$, $p > .1$, thus underscoring the absence of a sampling bias. This analysis was also conducted elsewhere in this paper and target naming was found to be similarly high and did not differ significantly by composite level, $p > .1$.

[6] Note that these results should not be taken to indicate that EvoFIT is in general better than sketch, since the composites are not directly comparable - good sketch artists typically do rather well (e.g. Frowd et al., 2005a).