Title: Identifying composites of famous faces: Investigating memory, language and system issues

Authors: N.A. Brace[1], G.E. Pike[1], P. Allen[2] and R.I. Kemp[3],

Addresses:

[1]Department of Psychology, The Open University, Milton Keynes, MK7 6AA, UK

[2] Department of Psychology, University of Westminster, London W1B 2UW

[3] School of Psychology, University of New South Wales, Sydney, 2052, Australia

Running title: Identifying composites of famous faces

Author responsible for correspondence:

Nicola Brace

N.A.Brace@open.ac.uk

Tel: +44 (0)1908 654525

Fax: +44 (0)1908 654488

Title

Identifying composites of famous faces: Investigating memory, language and system issues

Abstract

Previous research on composite systems has cast doubt on their potential to produce a good likeness. Poor memory of the target's face and problems in translating a verbal description into a visual mental image were examined as possible reasons for this. A computerised composite system was used to construct facial images of famous individuals. Results indicated that a reasonable number of composites were identified overall. Composites produced directly by the 'operators' tended to be more accurate than those produced via 'describers'. In addition, composites produced via describers, but not those directly by operators, were less accurate when created from memory, than when a photograph was provided. This suggests that verbal descriptions more than facial memory may limit composite accuracy.

Introduction

Composite systems are vital to police work as they allow trained officers to produce a likeness based on a witness' memory of a perpetrator's face. In the United Kingdom (U.K.) composites are usually constructed once an interview has generated a description, and their primary purpose is to generate a possible suspect and hence assist the police investigation. Early composite systems such as Identikit (which used line-drawn features and accessories) and Photo-FIT (which comprised actual greyscale images of facial features) were criticised for containing only limited feature databases and for using a piecemeal construction technique (e.g. Shepherd and Ellis, 1996). E-FIT, the compositing system currently used widely throughout the U.K., was developed by psychologists at Aberdeen University and was based on knowledge concerning face perception. This computerised system therefore has many potential advantages over the earlier systems. Importantly, the witness works with an operator to modify an image of the whole face. Global changes can be made to the facial image and there is greater flexibility in the placement of, and blending between, the different facial features. As well as a larger database of facial features, the system permits artistic enhancement via sophisticated drawing packages. The experiments reported in this paper used the E-FIT system.

Empirical work on earlier composite systems indicated that the resulting composites often bore little resemblance to the face they depicted. One important finding to emerge was that composite images constructed with the target present were found to be no better than those created from memory. Thus, removing the memorial component of the task

did not improve the quality of the composites constructed. Laughery and colleagues (Laughery, Duval and Fowler, 1977, and Laughery and Fowler, 1980) showed one of several different target faces to a pair of witnesses; one witness then constructed the face working with an Identikit technician and the other worked with a sketch artist. They found that the sketches were rated as more like the target than the Identikit constructions, and only the sketches showed a difference between those made from witness memory and those made with the target in view. Similarly, Ellis, Davies and Shepherd (1978) found that the accuracy of Photo-FITs created when the target was in view did not differ from those made from memory.

Recent evidence suggests that a better quality of composite image can be achieved with the more recent computer-driven systems. Cutler, Stocklein and Penrod (1988) found that participants could match composites produced by an experienced operator using the Mac-A-Mug Pro system to a photograph of the person the composite depicted with a success rate ranging from 58% to 80%. Wogalter and Marwitz (1991), using the same composite system, found that composite images produced while viewing the target face were significantly better than composite images produced from memory. Brace, Pike and Kemp (2000) showed pairs of composites of the same famous face created using the E-FIT system, one constructed from memory and one with a photograph in view, and participants were asked to choose which one best represented the famous person. The results showed that for composites created in the usual way (with a person describing the face to a composite operator) significantly more of those created with a photograph in view were chosen as the 'better' composite. Thus, removing the memorial component did improve the quality of the composites constructed using these computer-driven systems.

However, although computer-driven systems may have overcome some of the limitations of the earlier systems, other studies still cast doubt on the forensic value of composites constructed using these recent systems. Koehn and Fisher (1997) asked participants to interact with someone unfamiliar to them and then to construct the face of that person on the Mac-a-Mug Pro system with the help of an operator. Results revealed that likeness ratings given to the composites were poor and performance on a photo-spread task yielded a correct hit rate of only 40%. Using the same system, Kovera, Penrod, Pappas and Thill (1997) found that students were unable to identify a familiar person depicted in a composite image, even though these were of other students or faculty members known to them. Another finding was that composites constructed from memory using the modern E-FIT were no better than composites produced using the older Photo-FIT system (Davies, van der Willik and Morrison, 2000).

There are several different factors that may impinge upon the quality of the composites produced. An important consideration is the skill of the composite operator. Davies, Milne and Shepherd (1983) compared an experienced Photo-FIT operator (who had completed over 1000 composites) with a novice operator (who understood the mechanics of the kit but had little practical experience). Results suggested that the composites produced by the experienced operator were judged to be more like the target than those produced by the novice operator. In particular, the experienced operator was more skilled at the stage of feature selection. The importance of the skill of the operator was also highlighted in a study by Gibling and Bennett (1994) who asked six professional police Photo-FIT operators to produce one Photo-FIT using artistic enhancement and one without any artistic enhancement. A variety of methods of artistic enhancement were

possible so that, for example, unwanted areas could be blanked out and alterations could be made to the image in response to a request from the witness. Police recruits were asked to identify each target from a photographic line-up and results revealed that the composites that had been artistically enhanced were significantly more likely to be correctly identified than the composites that had not been enhanced. These findings suggest that when using Photo-FIT knowledge of and skills in enhancement techniques were important in producing a good likeness of the target face.

Other research has considered whether the poor quality of composite images could be linked to the initial verbal description given by the witness. For most composite systems, the first stage is to obtain a verbal description of the target face. The operator then constructs the initial composite image from this description. One exception is the Field Identification System where the witness or participant constructs the image without the aid of a second person. Using this system, Yount and Laughery (1982) asked ten participants to construct images of two familiar (e.g. roommate/husband) and two unfamiliar faces from memory. The unfamiliar faces were of targets seen live for around fifteen seconds. Results showed that whilst there was a difference in the time it took participants to construct images of familiar and unfamiliar faces, no differences emerged in terms of the quality of the images. Here quality was assessed by asking raters to judge the constructions on a six-point goodness-of-fit scale, when photographs of the target were projected onto a screen. The authors concluded that their results point to factors other than the verbal description as limiting the quality of the composites. If one assumes that the memory representation for a familiar face will be superior to that for an unfamiliar face, then the quality of the composites depicting the familiar faces should have been superior to those depicting the unfamiliar faces.

However, the usual method of constructing composites is based on the verbal information provided by a witness and this may impact on composite quality in two ways. First, there may be a problem with translating the verbal information into an accurate visual image. Christie and Ellis (1981) reported no statistically significant correlation between the accuracy of the verbal description provided and the accuracy of the subsequent Photo-FITs. In addition, they found that verbal descriptions emerged as more accurate than Photo-FITs in both an identification task and a sorting-by-target task. Secondly, asking witnesses to provide a verbal description could influence their ability both to recognise that face (as constructed in the composite) and to detect inaccuracies in the composite image. Schooler and Engstler-Schooler (1990) looked at the effect of verbally describing a face on a line-up identification task. They showed participants a video of a crime scenario and then asked one group to provide a verbal description of the perpetrator's face, a second to form a visual image of this face and a third to do nothing. Those who were asked to produce a verbal description were significantly less accurate in choosing the target from an array than the other two groups, who did not differ from one another. It was suggested that the verbalisation group was biased towards relying on the memory of their own description of the face rather than their initial visual memory of the face. That is, their verbal memory had over-shadowed their visual memory.

This verbal overshadowing effect has been replicated by Dodson, Johnson and Schooler (1997), who found that recognition was not only impaired when the participants had described the target face, but also when they had described a very familiar face or a different novel face. The authors proposed that verbalisation may be orienting participants to rely on feature information (which is relatively easy to verbalise) rather

than holistic/configural information (which is more difficult to verbalise). In support, Wogalter (1991) had found that asking participants to image the target's face produced significantly higher recognition performance in comparison to participants who had been asked to provide a verbal description using one of two different procedures. The two verbal description procedures themselves yielded different results; asking participants to generate their own descriptors resulted in better recognition performance than asking them to use a descriptor checklist. However, Yu and Geiselman (1993) failed to find any evidence of a verbal overshadowing effect when they compared recognition performance of three groups of participants; those asked to form an Identikit composite, those asked to write a verbal description and a control group.

There is therefore evidence that the quality of facial composites could be affected both by describing the face verbally and by the strength of the visual memory for the target face. The research reported in this paper explored the degree to which these two factors affected the quality of the resulting composites. The experiments reported here looked at the quality of composites produced by a computer-driven composite system in use today, and the impact of two variables on identification rates; namely whether the composites were constructed with target present or from memory and whether they were constructed directly by the operator or with the operator working with someone describing the face. The effectiveness of facial composites was assessed in a forensically realistic manner. As Morier (1995) and Bond and McConkey (1995) have pointed out, composites are used to trigger someone's memory, to allow someone who may know the perpetrator to identify the person portrayed in the composite. When the police publish a composite their hope is that the composite will be viewed and recognised by someone who is familiar with the perpetrator. It is therefore important that the quality of composites is assessed by judges

who are familiar with the target. To this end, participants were asked to provide likeness ratings or to identify composites of people they are likely to be familiar with, namely composites of famous faces.

Experiment 1

Method

Design

A 2 x 2 within-participants design was employed. The first factor was Mode of Construction: whether the composite was constructed with a describer or directly by the E-FIT operator without a describer. The second factor was Mode of Presentation: whether the composite was constructed from memory or with a photograph of the target present. Four dependent variables were employed in a rating task: whether the participant thought they would be able to identify the named target from a photograph, the participant's rating of the familiarity of the target, the participants rating of the distinctiveness of the target and how good a likeness the composite was of the target.

Participants

56 participants took part in the rating task as part of the requirements of their research methods course. All were undergraduate psychology students.

Composite construction

The targets were 48 famous people (who were all Caucasian males) selected by individuals not involved in constructing the composites. These famous people included British and American film and television actors and musicians of various ages, well-known sports personalities and one member of the British Royal family. Photographs

depicting these celebrities were obtained from a variety of sources. Two E-FIT operators were employed to create the facial composites. Each operator constructed two composites of half of the celebrities, 12 while working with a describer (filling the role of the witness) and 12 without. For each target the first composite was constructed from memory and the second with a photograph present. For those faces constructed with a describer, care was taken to keep the operator unaware as to the identity of the target. A breakdown of the construction stage is presented in Table 1.

Insert Table 1 here

The procedure used to create the composite with the describer followed that recommended to police E-FIT operators. First, a full verbal description was obtained from which the operator constructed the initial image. This was shown to the describer who then suggested changes to the face image. For each target, the first composite was constructed from memory and the second while looking at the photograph. Construction of composites in the memory and photograph present conditions was blocked, so that two composites of the same target were never created immediately one after the other. Both operators were trained in the use of the E-FIT system and had created composites before. However, to ensure that any increase in skill/experience during the construction stage did not disproportionately affect one condition, the order in which the composites from each condition were created was varied. Thus for each condition composites were created near the beginning, middle and end of the construction phase. See Figure 1 for examples of the composites constructed.

Insert Figure1 here

Materials

Materials consisted of the 96 composites of the 48 famous people and a response sheet. The 96 composites were divided into 4 sets of 24 composites. Each set comprised 12 composites constructed by the operator alone and 12 by a describer working with the operator. Of each of these, 6 had been constructed from memory and 6 with a photograph present. No famous person was depicted more than once in a set. The response sheet contained a table, with one row for each image and five columns. The first column contained the name of the person depicted in the image and brief additional information concerning their identity, e.g. 'Rowan Atkinson – British actor who plays Mr Bean and Blackadder'. The second column contained the question 'Did you know who this person was?' The remaining three columns contained a 10 point scale (where 1 was a low score and 10 a high score) for each of the following: how familiar participants were with the person; how distinctive a face did they thought the person had; and how similar was the composite to the person depicted (i.e. the accuracy of the likeness). To assist the participant, the response sheet contained two example responses, one for a face that was known and one for a face that was not known.

Procedure

Participants completed the rating task one at a time and were told that they could take as long as they liked to complete each rating. They were instructed to look at the name and additional information in the first column and to decide whether they knew this person to the extent that they would be able to recognise the person from a good photograph. If they decided that they did know the person, they were then asked to provide familiarity and distinctiveness ratings. These ratings would be based on their memory of the

person's facial appearance, as neither a photograph nor composite of the target's face was visible. Once these ratings were completed, participants were then shown the corresponding composite and asked to decide how similar it was to the person named in the first column. Participants were provided with one set of 24 images, placed in a pile so that only the relevant image would be visible at the appropriate time, and they were allowed to work through the set at their own pace. On completion participants were thanked and fully debriefed.

Results

Celebrity status of persons depicted in composites

Participants were asked to indicate whether or not they knew the celebrity whose face was being depicted in the composites. These data are presented in Table 2 according to condition, as participants rated different celebrities within each of these conditions. However, it is important to remember that the same people were depicted in the operator memory and photograph conditions and also in the describer memory and photograph conditions.

Insert Table 2 here

To assess whether the faces in the different conditions differed according to their celebrity status, a 2 x 2 within-participants ANOVA was performed. There was no statistically significant effect of either Mode of Construction ($F(1,55)=2.04$; $p=0.16$) or Mode of Presentation ($F(1,55)=0.00$; $p=1.00$ and no statistically significant interaction ($F(1,55)=0.76$; $p=0.39$).

Familiarity and distinctiveness ratings

For those famous people that were known to the participant, familiarity and distinctiveness ratings were collected via 10-point scales, where 10 was indicative of either very familiar or very distinctive. It should be noted that although no picture of the person in question was provided, the participant was asked to give ratings from their memory of the person's facial appearance and *not* from the composite, which remained face-down during this exercise. This was done to gain data on whether the faces in the different conditions differed according to familiarity or distinctiveness. The familiarity and distinctiveness data are displayed in Table 3.

As can be seen in Table 3 below, familiarity ratings were reasonably high, averaging around 6.9 on the 10-point scale. Distinctiveness ratings were lower, averaging around 5.8. Overall, there appears to be very little difference between the conditions for either the familiarity or distinctiveness data, although the mean ratings for the two operator conditions are slightly higher than those for the two describer conditions. These data were explored further using 2 x 2 within-participants ANOVA's. Analysis of the familiarity data revealed no statistically significant main effects of Mode of Construction [$F(1,55)=3.53$; $p=.066$], Mode of Presentation condition [$F(1,55)=0.70$; $p=.408$] or statistically significant interaction [$F(1,55)=0.19$; $p=.664$]. Analysis of the distinctiveness data revealed a statistically significant main effect of Mode of Construction [$F(1,55)=12.92$; $p<.005$], but no statistically significant main effect of Mode of Presentation [$F(1,55)=0.24$; $p=.625$] or statistically significant interaction [$F(1,55)=0.06$; $p=.806$].

Insert Table 3 here

In addition to the familiarity and distinctiveness ratings, the participants were also asked to rate the similarity of the composite to their memory of the target. Ratings were made using a 10-point rating scales, where 10 was indicative of 'very similar'. As can be seen in Table 3, mean similarity ratings were quite low, averaging around 3.6. The similarity data was subjected to a 2 x 2 within-participants ANOVA, which revealed statistically significant main effects of Mode of Construction [$F(1,54)=40.31$; $p<.001$], of Mode of Presentation [$F(1,54)=14.47$; $p<.001$] and a statistically significant interaction [$F(1,54)=6.49$; $p<.05$]. The ratings for the composites constructed by operator alone were higher than those for the composites constructed with a describer, and that whilst there appears to be little difference in ratings given to the operator composites constructed from memory and those with a photograph, those constructed by the describer with a photograph were given higher ratings than those constructed from memory.

Discussion

The data obtained in this first experiment revealed that overall just over 90% of the famous people were known by participants, who provided familiarity and distinctiveness ratings for the famous people they knew. There were no statistically significant differences in rated familiarity across the Mode of Construction conditions or the Mode of Presentation conditions; therefore familiarity was controlled for. A statistically significant difference in distinctiveness ratings only emerged across Mode of Construction conditions, suggesting that the faces the operators constructed were more distinctive than those constructed with the describers. However, the descriptive statistics suggest that this difference is in fact quite small, being less than half of a point on the 10-

point scale.

Participants were also asked to rate the quality of the likeness of the composite against their memory for the target using a 10-point scale. Analysis of these data revealed a statistically significant main effect of Mode of Construction and of Mode of Presentation as well as a statistically significant interaction. Higher similarity ratings were given to composites constructed by operators working alone than those constructed by operators working with a describer and, of the latter, higher ratings were given to those constructed with a photograph as opposed to those constructed from memory.

Similarity ratings are one measure of composite quality, however this measure may not accurately reflect whether the target depicted is identified, i.e. whether the composites of these famous faces will generate a name or other identity information. As this is the purpose of composites in police investigations, the aim of next experiment was to measure identification rates.

Experiment 2

Method

Design

The same 2 x 2 within-participants design was employed as in Experiment 1, with the two factors being Mode of Construction and Mode of Presentation. Two dependent variables were employed, these being identification accuracy and the confidence associated with the identification attempted. A correct identification response was one that involved the participant either correctly 'naming' the composite or providing specific information that would permit identification. Confidence was measured on an

11-point scale, with 0 marked as ‘guess’ and 10 ‘complete confidence’.

Participants

117 participants took part in the identification task as part of the requirements of their research methods module. All were undergraduate psychology students.

Materials

For the identification task, the 96 composites were divided into the same 4 sets of 24 composites as described in Experiment 1, with each set containing a balanced number of composites from the four conditions and never more than one composite depicting the same person. Each set of composites was copied onto overhead transparencies. Response sheets were drawn up so that participants could report the composite number shown, the information they could provide about the person depicted and their confidence score.

Procedure

Participants completed the identification task in groups of 14-18. They were told that they would be shown a series of slides and that on each slide there would be a composite, and that this was an image of a face that was constructed with the computer software the police use to reconstruct the faces of suspects. Participants were also told that in this experiment, these composites would be of famous people who should be familiar to them. Their task was to attempt to recognise them. Each composite would be displayed for 30 seconds, during which time they would need to write down 3 things: (1) the overhead letter and number, which would appear in the top right-hand corner, (2) the name of the person or sufficient information to identify them (an example was provided), and (3) a number between 0 and 10, which indicated how confident they were in their

decision. Participants were informed that if they were extremely confident that they knew who the composite portrayed then they should write down ‘10’, if they were extremely unconfident then to write down ‘0’, and to use the points in between accordingly. Participants were encouraged to make a response for each image, even if this was a complete guess, and that if their decision was a complete guess then they should remember to write down a figure of ‘0’ for their confidence.

Participants were asked not to talk to each other during the experiment and were given a practice trial in which they were shown a composite of Bruce Willis. They were reminded that if they could not remember the name, then they should write down information about the person that would identify him. They were told that for this example “the actor from Die Hard” or “Demi Moore’s ex-husband” would count as correct but “an American actor” would count as incorrect.

Each group of participants was shown one of the 4 sets of composites. Each composite was displayed for 30 seconds, during which time the participants recorded their responses and their confidence ratings using the forms provided. On completion, participants were thanked and fully debriefed.

Results

Scoring of responses

If a participant wrote the correct name or provided sufficient information to enable an identification to be made, then their response was scored as correct. If a participant wrote the wrong name or wrong/insufficient information to enable identification, then their response was scored as incorrect. It is important to note that using this scoring system, an

incorrect response is not equivalent to a false alarm.

Overall number of composites identified

Insert Figure 2 here

From Figure 2 it is apparent that 32 of the 96 composites were not identified by anyone, meaning that 64 (66%) of the composites *were* identified by at least one person. 21 composites were identified by up to 9% of the participants, 18 by between 10% and 29% of the participants, 12 by between 30% and 49%, 10 by between 50% and 79% of the participants and 3 by at least 80% of the participants. These 3 highly recognised composites were of Paul McCartney as created by one of the operators from memory, Rowan Atkinson (Mr Bean) as created with a describer from memory and Rowan Atkinson as created with a describer with the photograph present. See Figure 3 for examples of composites identified.

Insert Figure 3 here

Correct responses

Insert Figure 4 here

As can be seen from Figure 4, more composites were correctly identified when constructed directly by the operator than when constructed by a describer and operator together (the more realistic situation). The beneficial effect of working from a

photograph of the target is small in the operator construction condition; it is more marked in the describer construction condition.

Analysis of the data using a 2 x 2 within-participants ANOVA revealed a statistically significant main effect of Mode of Construction [$F(1,116)=26.28$, $p< 0.005$]. The main effect of Mode of Presentation failed to reach significance [$F(1,116)=3.18$, $p=0.077$], but the interaction between these two factors was statistically significant [$F(1,116)=8.01$, $p<0.005$].

Insert Table 4 here

Table 4 shows the percentage of correct responses broken down by whether the correct name or other correct information to permit identification was supplied. This table shows that approximately two thirds of the correct responses in each condition resulted from participants supplying the correct name.

Confidence Scores

Insert Table 5 here

Table 5 shows the confidence scores participants provided when making either a correct or incorrect identification. Confidence was higher when the response was correct than when incorrect.

These confidence scores were analysed using a 2 x 2 within-participants ANOVA. There was no statistically significant main effect of Mode of Construction [$F(1,41)=0.002$, $p=0.96$] or of Mode of Presentation [$F(1,41)=0.425$, $p=0.52$], but there was a statistically significant interaction between these two factors [$F(1,41)=11.005$, $p<0.005$]. Participants were slightly less confident when identifying a composite produced by a describer working from memory than one produced by a describer working with a photograph present. However, for composites constructed by operators working alone the opposite pattern emerged: participants were more confident in their identification of likenesses constructed from memory than directly from a photograph.

Incorrect responses

Insert Table 6 here

Table 6 presents the percentage of incorrect responses broken down by whether participants supplied an incorrect name or wrong/insufficient information about the celebrity. More incorrect responses were given when the composites were constructed with a describer than when constructed by the operator alone.

Analysis of the incorrect data, employing a 2 x 2 within-participants ANOVA, revealed that the main effect of Mode of Construction failed to reach statistical significance [$F(1,116)=3.39$, $p=0.068$]. There was no statistically significant main effect of Mode of Presentation [$F(1,116)=0.05$, $p=0.819$], and no statistically significant interaction [$F(1,116)=0.88$, $p=0.350$].

Relationships between familiarity, distinctiveness and similarity ratings and identification rates

Using the data from Experiment 1, it is possible to explore the relationships between the familiarity and distinctiveness ratings given to each target face and the percentage of people identifying that target face, and also to explore the relationship between the two different measures of composite quality, namely the similarity rating given to each composite and its identification rate.

Insert Table 7 here

Presented in Table 7 are Pearson correlation coefficients between the familiarity, distinctiveness, similarity ratings for each composite measured in Experiment 1 and the % of people that correctly identified that composite in Experiment 2. The similarity ratings significantly correlated with the number of times that composite was correctly identified in Experiment 2. Familiarity and distinctiveness ratings were also significantly correlated with the number of times composites were correctly identified, with the exception of those composites constructed by the operator from memory. Overall, the composites most often recognised in Experiment 2 were those of targets who were rated as most familiar and most distinctive.

Discussion

The data collected in this second experiment revealed that a substantial number of the targets depicted in the 96 composites were correctly identified. In fact, two-thirds of the composites constructed generated a name from at least one participant, which is perhaps

the most important statistic in terms of a police investigation (where only one person needs to generate the correct name). Given that Experiment 1 revealed that overall just over 90% of the famous people were known to a comparable sample of participants, the current data on identification rate is likely to be an underestimation; if the participant did not know the celebrity, then she or he would never have been able to identify the image.

Furthermore, analysis of the identification data measured in Experiment 2 revealed a similar pattern to the analysis of the similarity data in Experiment 1, and there was a significant correlation between these two measures of composite quality. The composites constructed directly by the operator were more likely to be correctly identified than the composites constructed by a describer working with an operator. There was also a significant interaction between Mode of Construction and Mode of Presentation suggesting that when a composite was constructed by an operator working alone, it made little difference to the composite quality whether he worked from memory or from a photograph. However, for composites constructed by a describer and operator working together, quality was significantly improved when the describer was working from a photograph compared to when he worked from memory.

Finally, there were statistically significant correlations between the familiarity and distinctiveness ratings given to each target face and the identification rates, except where the composites were constructed by the operator working from memory. Greater familiarity and distinctiveness of the face depicted in the composite were also associated with a higher rate of identification. As mentioned previously, participants' ratings of the similarity of the composite to their memory of the target were positively correlated with the identification rates measured in this experiment, suggesting that similarity ratings can

be a useful indicator of the quality a composite.

General Discussion

The findings reported here are generally positive regarding the potential of composites to trigger someone's memory and reveal the identity of the person depicted in the image. Of the ninety-six composites created, sixty-four were identified by at least one participant who saw these composites and ten were identified by at least half of the participants. It is also worth bearing in mind that for the police, publicizing a composite is often one of many avenues of investigation in serious crime cases, and having just one person provide a correct identification may greatly assist that investigation by generating a possible suspect. The figures reported here also underestimate the quality of the composites because they include the responses of participants who failed to identify the composite because they were unfamiliar with the celebrity being portrayed.

The present findings lend support to those of Wogalter and Marwitz (1991). In contrast to the findings of studies using earlier composite systems, such as Photo-FIT, the Wogalter and Marwitz study and the present study have shown that quality of composites created using the more recent computer-driven systems can be improved if a photograph of the target is present during the construction of the composite. However, in Experiments 1 and 2 reported here, this was only the case when composites were constructed with a describer. The fact that the quality of composites created by the operators was not significantly enhanced by the presence of a photograph suggests that rather than refreshing a mental image, the photograph is facilitating the verbal description provided by the describer. Significantly fewer composites created by a

describer were identified than those created by the operator working alone. This difference could well be due to the describer being required to verbally describe the face and to the operator then having to translate this verbal description into a visual image.

The findings presented here contrast with those of Kovera *et al*. (1997), who reported that only 1.7% of the names generated by composites of familiar targets were correct. Their composites were constructed from memory by a group of 'preparers' together with an experienced operator and given to other participants who should have known the persons depicted. This procedure compares closest with the 'describer from memory' condition reported here in Experiment 2, where the identification rate was 12.6%. The difference in results cannot be accounted for in terms of operator experience – in both studies an experienced operator created the composites. Also, in both, the composites were created by the operator working with someone who was familiar with the person to be depicted in the composite. Both are also comparable in terms of the additional information provided with the composite. In the Kovera *et al*. study, participants were told that the composites were of students and faculty members from high schools including their own, and in Experiment 2 participants were told that the composites depicted famous faces. It is possible that the difference in results is due to the stored representation of a famous face being somehow different to that of the face of someone with whom we are familiar with in real life. Another possibility is that participants are used to seeing 2D images, even caricatures and other non-veridical images, of famous people, but not of students and faculty members.

Although the data showed some variability in the extent to which these famous faces were rated as familiar, it is also possible that these famous faces were more familiar to

both the describers who worked with the operator to create the composite and to the participants attempting to identify the person depicted in the composite than were the faces of students and faculty members selected in the Kovera *et al*. study. Alternatively, the different findings may reflect the fact that in the Kovera *et al*. study, participants were only familiar with 20% of the targets (10 out of the 50) whereas in Experiment 2 it is likely that participants were familiar with approximately 90% of the celebrities. For example, seeing a large number of composites that they could identify may have prompted the participants in the present study to guess or to provide a response more often than the participants in the Kovera *et al*. study.

Another possibility is that the difference in identification rates observed between the results of the Kovera *et al*. study and Experiment 2 may reflect differences between the composite packages used. Kovera *et al*. used Mac-a-Mug pro which requires a more piecemeal construction of the facial composite than the E-FIT system used here. Whilst the former systems requires the witness to first select individual facial features which are then put together, with the latter system a witness provides a verbal description from which an 'initial face' is constructed. The witness then suggests modifications to this 'initial face' and thus can engage in configurational rather than (or as well as) a piecemeal processing.

Identification/naming data have been obtained in an earlier study using E-FIT by Davies *et al*. (2000). They compared E-FIT images with Photo-FIT images constructed from memory or with the target present, and the faces constructed were either familiar or unfamiliar to the participants who were creating the composites. Naming data indicated that when E-FIT composites of familiar faces were constructed with the target present,

nearly 50% of the composites were correctly named and when constructed from memory 17% of those depicted in the composites were identified. The results presented here are broadly in line with these figures - approximately 23% of the composites created by the operators were correctly identified. Unlike the Davies *et al*. study, the results of the experiments reported here did not show a statistically significant difference between identification rates for composites created from memory and those with a target photograph present. However, the operators who created the composites used in the experiments reported here had considerably more experience of using E-FIT than the participants in the Davies *et al*. study (both operators in this study had been trained and had already created composites previously). Furthermore, for the composites used in the experiments reported here, the operators used a specialized graphics package to refine the composites they created (a standard part of the E-FIT system), whereas the participants in the Davies *et al*. study did not.

In conclusion, the findings reported here indicate that the quality of composites is likely to be limited by the fact that the witness has to work together with a police operator to construct the composite. It appears that it is the verbalisation process rather than the quality of the visual memory for the target face that is one principal factor limiting the quality of the composite produced. Future research should therefore address how the operator might assist the witness to retrieve and describe the facial image and consider ways of allowing the witness to work directly with the computer-driven system itself.

Acknowledgements

This research was assisted by Nuffield Grant no. AT/100/97/0087. The authors would

like to thank Jim Turner for his comments on an earlier draft of the paper and also for acting as one of the describers. The authors would also like to thank Heather Dawson for acting as the other describer and Peter Bennett and Aspley U.K. Ltd for their help and advice.

References

Bond, N.W. and McConkey, K.M., (1995). Information retrieval: Reconstructing faces. In N. Brewer and C. Wilson (Eds.) *Psychology and Policing,* LEA.

Brace, N., Pike, G. and Kemp, R. (2000). Investigating E-FIT using famous faces. In A. Czerederecka, T. Jaśkiewicz-Obydzińska and J. Wójcikiewicz (Eds.), *Forensic Psychology and Law.* Kraków: Institute of Forensic Research Publishers.

Culter, B.L., Stocklein, C.J. and Penrod, S.D. (1988). Empirical examination of a computerized facial composite production system. *Forensic Reports, 1,* 207.218.

Davies, G.M., Milne, A. and Shepherd, J.W., (1983). Searching for operator skills in face composite reproduction. *Journal of Police Science and Administration, 11,* 405-409.

Davies, G., van der Willik, P. and Morrison, L.J. (2000). Facial composite production: A Comparison of mechanical and computer-driven systems. *Journal of Applied Psychology, 85, 119-124.*

Dodson, C.S., Johnson, M.K. and Schooler, J.W., (1997). The verbal overshadowing effect: why descriptions impair face recognition. *Memory and Cognition, 25,* 129-139.

Ellis, H.D., Davies, G.M. and Shepherd, J.W., (1978). A critical examination of the PhotoFIT system for recalling faces. *Ergonomics, 21,* 297-307.

Christie, D.F.M. and Ellis, H.D., (1981). PhotoFIT constructions versus verbal descriptions of faces. *Journal of Applied Psychology, 66,* 358-363.

Gibling, F. and Bennett, P. (1994). Artistic enhancement in the production of Photo-FIT likenesses: An examination of its effectiveness in leading to suspect identification. *Psychology, Crime and Law, 1,* 93-100.

Koehn, C.E. and Fisher, R.P. (1997). Constructing facial composites with the Mac-A-Mug Pro System. *Psychology, Crime and Law, 3,* 209-218.

Kovera, M.B., Penrod, S.D., Pappas, C. and Thill, D.L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology, 82,* 235-246.

Laughery, K.R., Duval, G.C. and Fowler, R.H. (1977). An analysis of procedures for generating facial images. *Mug File Project Report Number UHMUG-2.* University of Houston, Texas.

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

Morier, R.L., (1995). A study comparing various composite imagery techniques. *Journal of Forensic Identification, 45,* 381-395.

Schooler, J.W. and Engstler-Schooler, T.Y., (1990). Verbal overshadowing of visual

memories: Some things are best left unsaid. *Cognitive Psychology, 21,* 423-468.

Shepherd, J.W. and Ellis, H.D. (1966). Face recall – Methods and Problems. In S.L. Sporer, R.S. Malpass and G. Koehnken (Eds) *Psychological issues in eyewitness identification.* Lawrence Elbaum Associates (Hillsdale, NJ, England).

Wogalter, M.S. (1991). Effects of post-exposure description and imaging on subsequent face recognition performance. *Proceedings of the Human Factors Society, 35th Annual Meeting,* 575-579.

Wogalter, M.S. and Marwitz, D.B., (1991). Face composite construction: In-view and from-memory quality and improvement with practice. *Ergonomics, 34,* 459-468.

Yount, M.B. and Laughery, K.R. (1982). Facial memory: Constructing familiar and unfamiliar faces. *Bulletin of the Psychonomic-Society, 19,* 80-82.

Yu, C.J. and Geiselman, R.E., (1993). Effects of constructing Identi-kit composites on photospread identification performance. *Criminal Justice and Behavior, 20,* 280-292.

Table 1: The construction conditions of 96 composites of 48 famous faces

		Mode of Presentation condition	
Mode of Construction condition		From Memory	With Photograph
E-FIT Operator 1		12	12
E-FIT Operator 2		12	12
Describer 1	With Operator 1	6	6
	With Operator 2	6	6
Describer 2	With Operator 1	6	6
	With Operator 2	6	6

Table 2: Percentage of 'famous' people known to the participants by condition

Condition	**% Known**	
	Mean	***SD***
Operator memory	91.7	13.5
Operator photograph	90.5	14.5
Describer memory	88.7	16.6
Describer photograph	89.9	15.1

Table 3: Familiarity, distinctiveness and similarity ratings

Condition	**Familiarity**		**Distinctiveness**		**Similarity**	
	Mean	***SD***	**Mean**	***SD***	**Mean**	**SD**
Operator memory	7.0	1.5	5.9	1.5	3.9	2.5
Operator photograph	7.1	1.5	6.0	1.3	4.0	2.5
Describer memory	6.8	1.6	5.6	1.5	2.9	2.1
Describer photograph	6.9	1.5	5.7	1.4	3.6	2.3

Table 4: Type of identifying information supplied

Condition	Name		Information	
	Mean %	***SD***	**Mean %**	***SD***
Operator memory	19.5	*18.0*	4.0	*7.8*
Operator photograph	18.8	*17.7*	3.6	*8.1*
Describer memory	12.5	*9.5*	0.1	*1.5*
Describer photograph	17.2	*13.8*	0.9	*3.7*

Table 5: Confidence scores (/10: 0 = complete guess and 10 = complete confidence)

Condition	Correct		Incorrect	
	Mean	**SD**	**Mean**	**SD**
Operator memory	6.9	2.9	2.9	2.7
Operator photograph	6.1	3.0	2.9	2.6
Describer memory	6.9	3.4	2.7	2.5
Describer photograph	7.0	3.1	3.4	2.8

Table 6: Type of incorrect identifying information supplied

Condition	**Name**		**Information**	
	Mean	***SD***	**Mean**	***SD***
Operator memory	23.4	*21.3*	9.5	*17.1*
Operator photograph	22.7	*21.5*	11.4	*15.9*
Describer memory	27.1	*24.0*	10.5	*15.1*
Describer photograph	24.8	*23.5*	11.0	*18.3*

Table 7: Correlation coefficients (rating scores correlated with % identification rate)

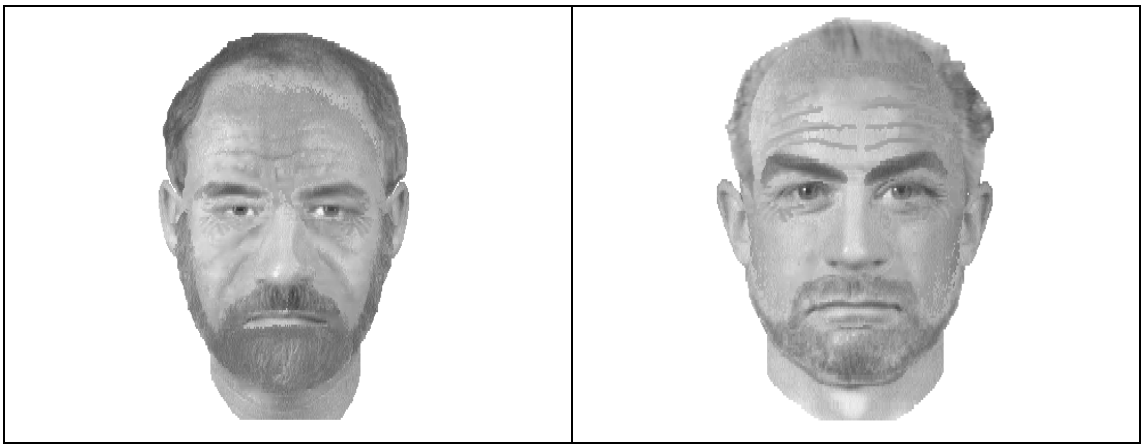
Condition	Familiarity	Distinctiveness	Similarity
Operator memory	.24	.24	.80**
Operator photograph	.47*	.63**	.56**
Describer memory	.61**	.65**	.85**
Describer photograph	.69**	.82**	.61**
All conditions	.52**	.60**	.70**

* sig. at $p < 0.05$ level ** sig. at $p < 0.01$ level 'N' for all cells is 96

Figure 1

Examples of the composites constructed

Composites of Sean Connery constructed by the operator working alone from memory (left) and from photograph (right).



Composites of Prince Charles constructed by an operator with a describer working from memory (left) and from photograph (right)

Figure 2: The number of composites correctly identified

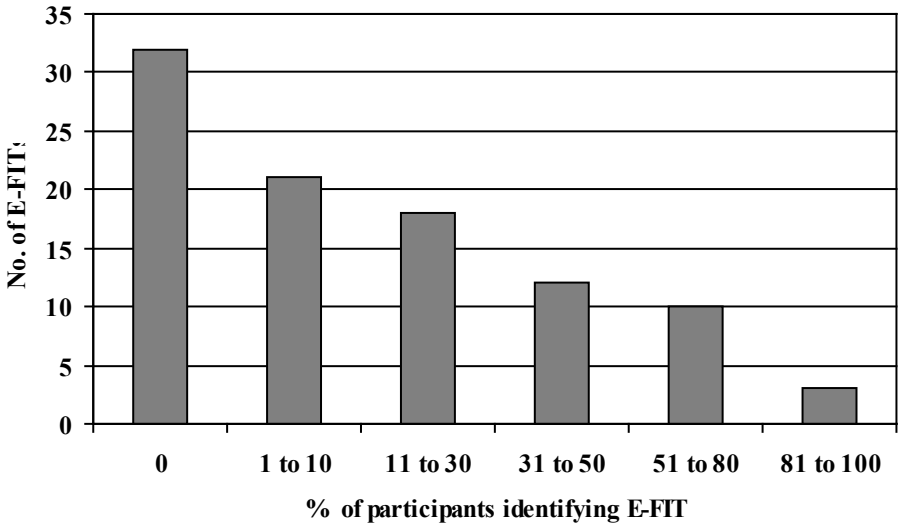

35
30
25
20
15
10
5
0
No. of E-FITs
0
1 to 10
11 to 30
31 to 50
51 to 80
81 to 100
% of participants identifying E-FIT

Figure 3

Examples of composites identified

Paul McCartney: composite constructed by operator working alone from memory. Identified by 97% of participants. 	Jack Nicholson: composite constructed by operator working alone from photograph. Identified by 61% of participants. 
Bill Clinton: composite constructed by operator with describer working from a photograph. Identified by 37% participants. 	Robbie Coltrane: composite constructed by operator with describer working from a photograph. Identified by 6% of participants. 

Figure 4: Mean percentage correct responses by condition