The Relative Utility of Verbal Descriptions and Facial Composites in Facial Identifications

Agnieszka M. Lech, Robert Johnston

University of Kent
agnes.m.lech@gmail.com, R.Johnston@kent.ac.uk

Abstract

Research on early facial composite systems has often demonstrated their poor capability to produce a good likeness of a target face. Photofit composites have been shown to produce poorer identification than simply furnishing a description of a target face. The current study revisits this comparison with composites produced using a holistic based composite system, EFIT-V. The effectiveness of verbal descriptions of faces was compared with facial composites in two laboratory tasks. Prior to the empirical work, eight 'witnesses' viewed an unfamiliar face and then provided a verbal description and created a facial composite. In Experiment 1 participants evaluated the relationship between a target face and three combinations of witness information (description alone, composite alone and description and composite combined). Providing a description was rated as more useful than providing a composite. In addition, presenting both a composite and description together was rated as less useful than presenting a description on its own. In Experiment 2 participants were supplied with the same types of information while they attempted to choose the target from an array of faces. Better performance was achieved from a description than from a composite or a description and composite combined. Analysis of participants' confidence in their decisions showed it was higher when prompted by descriptions, irrespective of whether the decisions were correct or not. The implications of these findings for the utility of facial composites are discussed.

Keywords: *Photofit composites, EFIT-V, facial composites, verbal descriptions*

1. Introduction

The judicial system is provided with a valuable source of information when an eyewitness can recognize an offender's face from presented photographs or a line-up. However, in the absence of a suspect and when an eyewitness cannot identify the offender from collections of mug shots, then constructing a facial likeness might be a quite useful way to seek the identity of the offender. Currently, there are a number of composite systems which were designed and developed in order to enhance witnesses' accuracy in creating a facial likeness of the offender.

Unfortunately, although humans are expert at recognising familiar faces their ability to process unfamiliar faces is much poorer [1]. Few people are likely to be capable of producing a highly accurate representation of a previously unfamiliar face. Especially one that might have been viewed only briefly and perhaps under a situation of stress. Facial composites are intended to help police in their investigations by generating the facial appearance of potential suspect. In real life, a witness/victim will create a facial composite of a person unfamiliar to

them and it is hoped that this likeness - even if impoverished - will be identifiable by someone who is already familiar with that person [2].

However, studies analysing the effectiveness of feature based systems like Indekit or Photo-fit in majority cases indicate rather low level of accuracy [3]. Both systems require eyewitness to create a face from number of various features components, which is a difficult task and prone to mistakes since faces are said to be stored holistically in memory. Those results gave the beginning to development of composite systems where witnesses were working on a holistic facial composition [3] because judgements of features are more accurate when they are based on the whole face rather than in isolation. Systems based on evolutionary procedures are design in order to enhance the recognition which is found to be easier then recalling.

These new systems are under continuing development and research is needed to establish their effectiveness, however, Frowd et al. [4] conducted a comparison of 5 systems where 3 different evaluation techniques were used. Composite naming yield very poor results across all the systems, especially in a case of two leading UK systems PROfit and E-Fit; which composites were named correctly only twice in 300 attempts [5]. EvoFITs composites were named significantly better than combined PROfit and E-Fit composites, yet still Sketches were the best technique to create composites with highest rates of naming. In Frowd et al. study [4] participants were asked to match composites with target photographs, this time sketches and E-FITs gained the highest correct scores. In addition, results gained in laboratory studies, with different level of ecological validity, indicate that composites created in systems like E-Fit, PROfit, Faces and Identikit 2000 are named correctly about 20% of the time [6], followed by 10% for sketches, 5% and 3.5% for composites created in Photofit and EvoFit respectively [5].

Unfortunately, empirical investigation of existing composite systems has led to disappointing conclusions; composites are often evaluated as being of poor quality. Constructed faces are often of poor likeness and can be hard to match to the target face [7]. This has led some researchers to claim that a verbal description would be a better option for an eyewitness than a visual likeness [7]. Most police officers believe that descriptions provided by witnesses are an important lead in a criminal investigation. They are generally judged to be incomplete but still accurate [8]. The majority of facial composite systems are based on the verbal description provided by a witness. It is the police's standard procedure for the operator of the system to elicit a verbal description of the target face from the witness.

Three decades ago Hadyn Ellis succinctly characterised the nature of the problem, "The witness has some internal representation of the facial appearance of the person, but by what means can this be successfully communicated to others?" [9]. Research on early facial composite systems has often demonstrated their poor capability to produce a good likeness of a target face. In a recent and extensive review of the use of facial composites, Davies and Valentine [10] reach a pessimistic conclusion regarding their utility, but allow that the fourth generation composite systems that make use of a genetic algorithm have yet to be completely tested. This study continues that process of evaluation. One failing of the earlier composite systems has been demonstrated by Ellis and Christie [9]. In a number of identification tasks they found that providing facial composites led to poorer results than simply furnishing a description. The current study is intended to revisit this earlier study but examines the effectiveness of composites produced using one of the most contemporary fourth generation systems (EFIT-V).

Christie and Ellis [9] compared descriptions and composites by means of an identification task and a sorting task. In the identification task participants were provided with either composites or descriptions and asked to identify six target faces from an array of 24 faces.

The descriptions were significantly more effective in securing identifications than the composites. In the sorting task participants were given the six target faces and either 36 descriptions or 36 composites. They were asked to sort the descriptions (or composites) into groups that were associated with particular targets. Once again performance was shown to be markedly better when using the descriptions. As a post hoc addition to their main study Christie and Ellis [9] tested further participants on their identification test by giving them a combination of description and composite but found that performance was no better than had been achieved by participants in the main study who had received descriptions on their own. Indeed, descriptions on their own and descriptions and composites combined yielded mean scores out of 10 of 4.8 and 4.4, respectively.

Overall, Christie and Ellis [9] construct a strong argument for a composite simply being less effective than a description. As part of their post hoc study they asked each participant who had used the descriptions and composites in combination to evaluate which component was more useful to them in the identification task. This revealed a clear preference for descriptions over composites: 35 participants preferred the descriptions, 15 preferred the composites and 10 expressed no preference. They propose that the effectiveness of Identikit and Photofit composites might simply arise through the operation of these systems. "The feature-by-feature approach to constructing a likeness of a face shared by the two systems might be entirely inappropriate if the face is perceived and stored as a gestalt" [9].

If that argument is correct then the output of contemporary composite systems might fare better than the composites employed by Christie and Ellis. Most recently, advances in computers have allowed the creation of sophisticated software systems (e.g., E-FitV: [11]; EvoFIT: [12]), where the construction of composites is based on holistic processing rather than feature one. The newest generation of composite systems is based on evolutionary procedures that allow witnesses to select images that are most similar to the target. A witness starts with an array of faces from which they choose a face similar to the target. The composite system uses each chosen face to create ("breed") a new set of composites relatively similar to the "parent" face. This procedure is repeated until the witness cannot choose between the faces, which at this point should equally resemble the target or the system has failed to reached the desired face [10].

This current study has adopted a version of a task used by Christie and Ellis in their post hoc task it we sought their direct evaluation of different sorts of identification information produced by the witnesses. They were asked to make a series of ratings about the quality of the identification information provided in conjunction with the target faces. Participants were informed that the identity information had been reconstructed from memory by an earlier witness who had viewed the target briefly. They were instructed to make judgements about the utility of this information. Specifically, participants were asked to rate the identification information in terms of its likelihood of leading to the identification of this target and participants' confidence in their decisions. In Experiment 2 the same identification information (description, composite or description plus composite) was presented to participants who were asked to use it as a basis to select the target face from an accompanying line-up of six faces.

2. Experiment 1.

An ideal laboratory analogue of evaluating how well facial composites function should operate in the following way. Witness participants should briefly view an unfamiliar target face and then construct a composite from memory. This composite should be shown to a second group of participants who are already familiar with the target to see if they can

identify him/her. Few studies have managed to reconstruct this situation [13] but others adopted procedures that give only an approximation to the real world set up. The idea that witnesses construct composites of unfamiliar faces, which are later identified by participants, is considered to be a golden standard in psychological research. However, it is also possible for police to ask the public whether they have seen a suspect in the surrounding area. Therefore, the person making a potential identification does not necessarily have to be familiar with the suspect. He/she might recall seeing someone matching the description/composite in a restaurant or shop a couple of minutes earlier.

The current study attempts to extend the previous investigations to address composites produced via a holistic approach that may be more compatible with the witnesses' internal representation of the face they have seen. We also attempt to make the experience of the composite constructors more like real life witnesses as was recommended by Christie and Ellis. The composite constructors were invited in on one occasion to view the unfamiliar targets and again, three days later, when they attempt to produce a description and composite. This will introduce a more realistic time delay than just few minutes difference between seeing a face and creation of the composite. In the Christie and Ellis study the relative utility of combining descriptions and composites was tested in a post hoc way but we intend to examine this directly. In the real world, comparatively few of the composites created by witnesses are ever circulated. For example, police forces have not been keen to circulate composites when the confidence of the creator is low (Davies & Valentine [10] report that fewer than 10% are released). One imagines such a practise is largely based on commonsense assumptions, especially when one considers the low effectiveness of even superior composites. However, it is also possible to make a case that more information is always better than less information. By directly evaluating the effectiveness of descriptions and composites combined we may find evidence to either support or modify current police practice.

3. Composite Construction Phase.

Eight volunteers (4 Males and 4 Females; 4 Black and 4 White participants) acting as 'witnesses' were presented with two previously unknown target faces on a projector screen one at a time. Targets photographs were obtained from Physics Department (University of Kent) facial database. Each face was presented on white background for a period of 12 seconds. The targets presented to participants were 10x15 centimetres photographs which were viewed at a distance of approximately 0.5 m.

Each participant returned to the laboratory three days later and took part in the composite construction procedure. Participants first related a description of the target. Free recall was used to elicit these descriptions because it is usually the first step in an interrogation procedure. Therefore the eyewitnesses were asked to simply describe what they remembered about the target. Immediately afterwards they constructed composites using a computerised composite system called EFIT-V. Participants created their composites while assisted by an operator trained in the use of this system. The order in which they produced composites for black and white faces was counterbalanced across participants. This furnished the 8 descriptions and 8 composites that were used in the second part of the study.

EFITV (See Figure 1) is newly designed program which utilizes a multistage–evolutionary process.

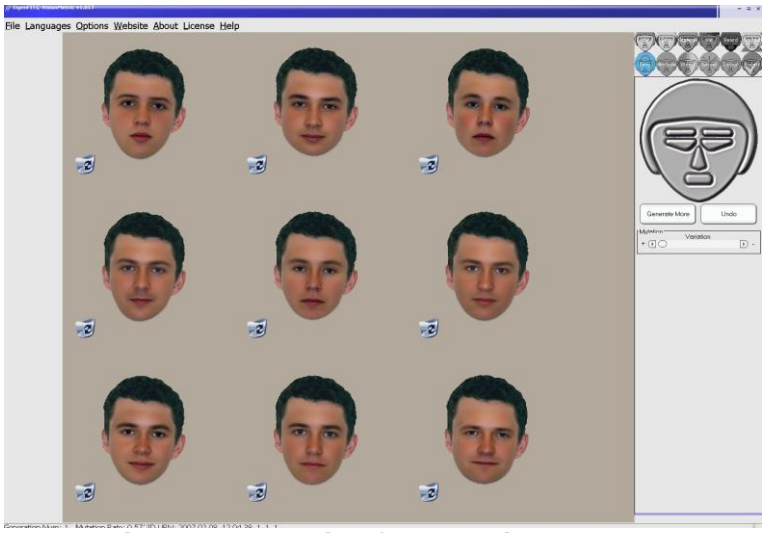

Figure 1 Example of EFITV View Screen

One starts working with EFITV by selecting the ethnicity (White, Black, and Asian) and gender of the subject. The second step involves choosing offender's hairstyle; the witness can mark, in any order, colour, length, style and type of hair. The witness is presented with a screen of nine different hairstyles that matched the best the chosen criteria. At each stage, an eyewitness is presented with nine different faces (computer synthesised face-images that are known as a generation). All that is required from an eyewitness is to choose the "face-image" from the selection, which shows the best likeness to the offender. In an evolutionary way, the chosen face is then cloned and mutated by the software in order to produce nine new faces. The witness repeats this procedure until he/she reaches the computer generated face that is the most similar to their memory of offender's face.

There are several options available in the software to ease the construction of the portrait. All features can be scaled, moved and rotated according to witness' opinion. In addition, features like ears, beards as well as glasses and hats can be added and adjusted.

Witnesses can change the face shape as well as the shape, orientation and location of the face's inner features. Again, the operator can rotate, scale, and move these features, both individually (e.g., only nose) or as a group (e.g., nose and eyes). This option helps to give the constructed face more individual appearance. Other possible functions let witness to make generated faces look older or younger as well as darker or lighter. It is also possible to mix faces from nine generated faces presented on the screen, in other words, if witness declares that faces number 2, 5 and 7 are somewhat similar to the face he/she had seen, one could create new generation of faces by mixing those three faces proportionally to witness' specifications.

4. Method

4.1. Participants

Participants were recruited by means of a link on a psychological research webpage. Because only a small number of respondents came from Black or other ethnic groups only the responses from White participants were processed further. The data from 94 White participants were entered into the analyses.

4.2. Materials

The experimental stimuli comprised the two target pictures used to elicit descriptions and composites in the composite creation phase. These images were presented in conjunction with either descriptions only, the composites only or both pieces of information simultaneously.

4.3. Design

The experiment employed a 2 Race of target (Black or White; repeated measure) x 2 Compositor race (Black or White; repeated measure) x 3 Identification information (Description only, Composite only, or Description plus Composite; repeated measure) design. The dependent variables were evaluations of the quality of Identification information in the form of ratings of Likelihood of identification based on the provided information on a 5-point scale (1 = very unlikely, 5 = very likely). Ratings of confidence for the identification information were also collected on a 5-point scale (1 = very unconfident, 5 = very confident).

4.4. Procedure

All participants were presented with each of the identification conditions (descriptions, composites, and descriptions plus composites) and asked to make the three rating decisions. First, all participants were presented with the descriptions and target faces, followed by the sequences of composites and photograph of the same targets. The order of the type of identification information was constant across all the experimental trials. However, the order of presentation of the items was randomised so each participant had a different order of the identification sequences.

There were 8 identification trials based solely on descriptions and 8 based solely on composites. In addition there were 8 identification trials where participants were asked to compare both the composite and given description with the target photographs, which together gave 24 identification trials. In all identification conditions participants were asked to compare the similarity between the provided information (either description, composite or description plus composite) and a target photograph. Analogously to the public being asked whether they have seen a person matching the identification information, in the current study participants were asked to make three evaluations for each target face and information pairing. They were asked to judge the Likelihood of identification based on the provided information and to rate how confident they were that the identity information related to this particular target.

5. Results

5.1. Analysis of ratings

A 2 x 2 x 3 ANOVA conducted to examine the effects of Target Race (White or Black, repeated measure), Creator Race (White or Black, repeated measure) and Identification Information (Description or Composite or Description&Composites, repeated measure).

5.2. Analysis of Likelihood

Table 1 shows the mean ratings of Likelihood distributed across Race of target face and Compositor race.

Table 1 Mean ratings of Likelihood.

White Targets						Black Targets					
White Compositor			Black Compositor			White Compositor			Black Compositor		
Desc	Comp	D&C	Desc	Comp	D&C	Desc	Comp	D&C	Desc	Comp	D&C
3.19	2.21	2.7	2.83	1.93	2.16	2.97	1.92	2.53	2.53	1.81	2.36

A significant main effect was found for the Target Race $F(1,93)=18.76$, $p<.01$, indicating higher Likelihood of identification for White composites (M=2.50) than for Black ones (M=2.35); for the Creator Race $F(1,93)=81.70$, $p<.01$ suggesting that composites constructed by White participants (M=2.59) scored higher on the Likelihood rating than those constructed by Black participants (M=2.27); and for the Identification Information, $F(2, 186)=91,83$, $p<.01$. These differences were tested using pairwise comparisons, adjusted for the family of significant differences at the .05 level; Bonferroni correction was applied. Similarly to previous findings all three means were significantly different from each other, mean Likelihood was the highest for the Descriptions (***M***=2.88), then for the Descriptions&Composites (***M***=2.44) followed by Composites (M=1.96).

All interactions were found to be significant. An interaction between Identification Information X Target Race $F(2,186)=9.76$, $p<.01$, suggested that White composites had higher Likelihood of identification than Black targets, based on the information provided in Descriptions (M=3.01 and M=2.75 respectively) and based on the Composites (M=2.07 and M=1.86 respectively). The difference between White and Black targets disappeared when Likelihood of identification was based on both Descriptions and Composites (M=2.43 for White and M=2.44 for Black) significantly lowered the Likelihood of identification. All together, Descriptions were found to be significantly different from Descriptions&Composites and Composites alone.

A significant Identification Information X Creator Race interaction $F(2,186)=5.09$, $p<.01$ indicated that composites created by White participants were rated higher on the Likelihood scale above all Identification conditions than those created by Black individuals. Again, the Descriptions (M=3.08 and M=2.68 respectively) were found to be the most effective source of Identification Information, followed by Descriptions&Composites (M=2.61 and M=2.26 respectively) and Composites (M=2.06 and M=1.87 respectively).

There was a significant Target Race X Creator Race interaction $F(1, 93)=6.34$, $p<.05$, with White composites created reaching higher level of identification Likelihood than Black ones. Post hoc analysis also supported the significant difference between Likelihood of identification of White (M=2.7) and Black (M=2.47) targets based on the information provided by White participants. The difference did not reach significant level for the information provided by Black participants (M=2.3 and M=2.23 respectively).

However, all these effects and interactions were modified by a significant Target Race X Creator Race X Identification Information interaction, $F(2,186)=6.90$, $p<.01$ (See Figure 1) was found. The only nonsignificant difference between Identification Information condition (specifically between Descriptions and Descriptions & Composites) was obtained for Black targets created by Black participants (M=2.53 for Descriptions and M=2.36 for the Descriptions&Composites). The White participants seem to provide information with higher Likelihood of identification than Black participants. It was also observed that White targets were reaching higher Likelihood than Black composites for White and Black participants across Identification Information conditions. The only exception was found for the White (M=1.93) and Black (M=1.81) composites created by Black participants.

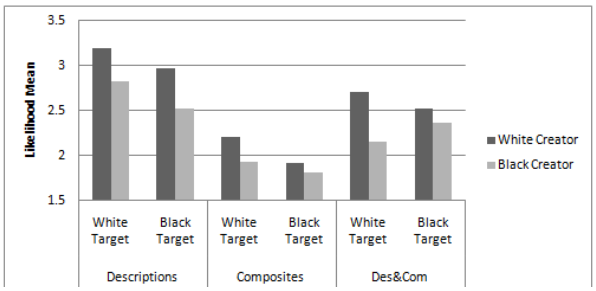

Figure 2 Relationship between Identification Information, Target Race and Creator Race for Likelihood rating

5.3. Analysis of Confidence

The main effect of Target Race was found to be significant $F(1,93)=4.02$, $p<.05$, with White targets ($M=3.7$) reaching higher level of confidence than Black targets ($M=3.64$). Identification Information X Creator Race $F(2,186)=4.87$, $p<.01$ was the only significant interaction (See Figure 2). Bonferroni pairwise comparison indicated that participants declared higher confidence for Descriptions provided by White ($M=3.76$) rather than Black ($M=3.62$) participants. All other differences did not reach the significant level.

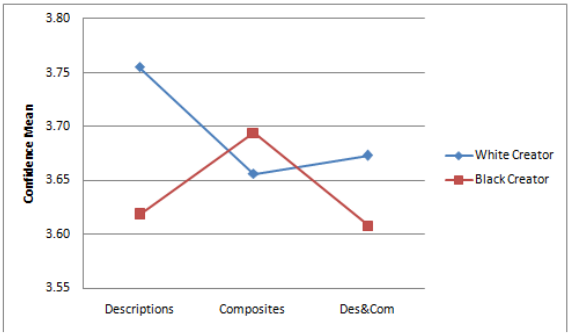

Figure 3 Relationship between Identification Information and Creator Race for Confidence rating

All other main effects and interactions were found to be non-significant.

6. Discussion

In many cases, when an offender cannot be recognised directly by an eyewitness, the criminal justice system needs to rely on a third party identification of a suspect based on a composite and/or a description. This study attempted to evaluate the relative utility of descriptions and composites created in EFITV. It required participants to consider pairings of target faces and accompanying identification information and evaluate the quality of the latter. It is noteworthy that there was an impact of the race of the compositor in that descriptions and composites constructed by White witnesses were evaluated as more useful than information elicited from Black witnesses. Given that all our participants in Experiment 1 were White we assume this reflects a shared frame of reference of the white compositors and White participants that leads them to emphasise the same sort of face information.

However, as our study was not able to recruit substantial numbers of black participants we will not attempt to comment further on this finding.

The finding of principal interest was when the accompanying identification information was a description alone this was judged by participants to be of higher quality than if the identification information took the form of a facial composite. Despite the fact that holistic software was used in order to create composites, it was still found that descriptions seemed to be the dominant Identification Information and they seemed to significantly enhance the likelihood of identification. Furthermore, the quality of the description alone was judged to be higher than that provided when the description was provided in conjunction with the composite. Finally, participants were also more confident when making identifications based on Descriptions rather than other information provided.

Christie and Ellis [9] found that a verbal description was more useful in securing an identification of a target than a composite produced using the Photofit system. However, composites constructed using Photofit are based on a feature-by-feature approach to creating composites that might be incompatible with how the memory of the face is represented. This experiment employed composites constructed with the EFIT-V system which is based on holistic-approach to composites' construction. Nevertheless, our findings mirror those of the earlier study. The descriptions alone were judged more likely to lead to correct identification then the composites alone.

Christie and Ellis [9] suggested that the superiority of descriptions is due to a more efficient use of processing resources. In a situation of creating a composite the witness needs to shift his visual memory to the "act of making up the composite" but this shift is not required when the eyewitness is asked to verbally describe the face. Christie [9] argued that descriptions were still more likely to lead to correct identification than Photofit composites even when they were constructed with a target face present.

In addition, when a person is asked to match unfamiliar faces this process relies on image matching rather than on a more sophisticated face matching strategy [14]. In such case a likeness matching procedure is a measure of individual feature likeness [15]. Thus, participants might expect the composite to be perfectly identical with the face of offender/target, which might lead to lower accuracy results. This might be a reason why participants/eyewitness will never reach the highest level of satisfaction from their composites. It is because they simply expect the composite to be identical with the face of offender/target, yet one can only reach only "as good likeness as possible" level. As McQuiston-Surrett et al. [16] stated "composites are only designed to reflect a person's physical appearance in a general sense in order to narrow down the pool of potential suspects".

Furthermore, in recognition of the face, both the spatial relations among the facial features as well as the shapes and sizes of the individual feature play an important role [17]. It was found that recognition of a face can be significantly disrupted when the spatial differences between the features are changed even slightly [17]. However, in general, people have problems with recalling and describing the individual facial features as well as spatial information among them [8], partially because we perceive faces in holistic matter rather than as a collection of features. This could lead to the impairment of created composites and influence the results gained from the rating tasks [15], hence participants might have compared the likeness between individual features rather than a general likeness between composite and target picture.

One might argue that reliability of this study was limited by the fact that only one White and one Black face were introduced to participants as a target face. In most cognitive and forensic experiments participants are asked to identify many more faces. However, as noted

by previous researchers such situations are lacking in ecological validity [18]. One can also argue that the use of photographs might be a limitation of current research. However, it has been shown that the real faces and photographs may be regarded as equivalent [9].

These findings appear to support the advantages of descriptions over composites when evaluated using judgements of usefulness in securing identification. Participants' ratings of the descriptions and composites produced using EFIT-V show a similar pattern to that previously seen with Photofit composites. In order to examine this question in a different way, a second experiment was conducted where the value of the identification information was tested more directly. In this case, participants were required to attempt to identify a face from a photo array using the different types of identification information. Nevertheless, it would be valuable to also have a more direct test of their relative effectiveness based on an identification task.

7. Experiment 2

In Experiment 1 the effectiveness of identification information that had been elicited from 'witnesses' was evaluated by asking participants to simply rate how useful they thought it would be in leading to the identification of the target it was intended to describe. The objective of Experiment 2 was to compare identification information using an identification task where the accuracy of participants' response could be measured. This was intended to mimic a task closer to the real life application of such materials. Participants were shown some identification information (description, composite or description and composite) in conjunction with an array of six faces. They were asked to select the target that was the best match to the presented identification information. Based on findings from Experiment 1, it is expected that Descriptions again will reach higher level of participants' ratings. However, it is predicted that Composites and Descriptions & Composites conditions should lead to higher level of Accuracy than identifications based solely on provided Descriptions.

8. Method

8.1. Participants

Thirty-four participants completed this on-line study. However, due to the small number of participants from Black and Other ethnic groups (4 participants) only the data from the thirty White participants (21 females and 9 males) was analysed. The mean age of participants contributing data was 27 years (ages ranged from 18– 53 years of age).

8.2. Design

The design of the experiment 1 did not allow analysis of accuracy, which was addressed in this experiment. The main aim of the current study was to examine the identification accuracy and confidence based on the different types of Identification Information for White or Black targets provided by White or Black witnesses.

The experiment employed a 2 Target Race (Black or White; repeated measure) x 2 Compositor Race (Black or White; repeated measure) x 3 Identification information (Description only, Composite only, or Description plus Composite, repeated measure) design. The analysis was conducted separately for Correct Identifications (Hits) and False Identifications (FI). The dependent variables were evaluations of the quality of Identification information in the form of ratings of Likelihood of identification based on the provided

information on a 5-point scale (1 = very unlikely, 5 = very likely) and ratings of confidence for these decisions were also collected.

8.3. Materials

The descriptions and composites created for the purpose of the previous study were also used in Experiment 2. In addition, arrays of six faces were constructed from combinations of the target face and five distractor faces. Targets photographs were obtained from Physics Department (University of Kent) facial database. The targets presented to participants were 10x15 centimetres photographs which were viewed at a distance of approximately 0.5 m. All White faces were presenting white males on the white background with blond, spiky hair. Black targets were also presented on a white background, all had short black hair. All faces had neutral facial expressions.

8.4. Procedure

The study was designed and conducted as an online study, posted on www.facebook.com, www.onlinepsychresearch.co.uk, www.psychologyonline.uk and http://psych.hanover.edu/research/exponnet.html.

Participants were shown each type of identification information (Descriptions, Composites and Descriptions & Composites) and asked to choose the most appropriate target face from the array of six faces. The target face was always present in the array (i.e., it was always a target present condition). After making their choice they were asked to rate their confidence in their decision on a 5-point scale (1 = very unconfident, 5 = very confident).

First, all participants were presented with the descriptions and target faces, followed by the sequences of composites and photograph of the same targets. The order of the type of identification information was constant across all the experimental trials. However, the order of presentation of the items was randomised so each participant had a different order of the identification sequences. As previously, there were 24 identification trials.

9. Results

9.1. Analysis of Accuracy

Participants had to choose the target face from the line-ups by using the provided information. There were six members in the line-up and it was always a target present condition. This gave the researcher two possible dependent variables – Hit (Correct Identification) and False Identifications (FI). The number of occasions when the participant correctly identified the target face in the array was counted, this is shown in Table 2. The mean level of detection across all conditions was approximately 20%.

Table 2. Mean number of successful attempts to identify the target in array of faces.

White Targets						Black Targets					
White Compositor			Black Compositor			White Compositor			Black Compositor		
Desc	Comp	D&C	Desc	Comp	D&C	Desc	Comp	D&C	Desc	Comp	D&C
1.33	.97	1.07	.77	.23	.63	1.13	.87	1.23	.57	.47	.53

A 2 x 2 x 3 ANOVA was conducted to examine the effects of Target Race (White or Black, repeated), Compositor Race (White or Black, repeated) and Identification Information (Description or Composite or Description & Composites, repeated).

There was a significant main effect of Creator Race, $F(1,29)=36.81$, $p < .001$. Participants were more successful with Identity information supplied by white compositors than by black compositors (mean identification accuracy was 1.1 and 0.53 respectively. There was also a significant main effect of Identity information, $F(2,58) = 5.04$, $p < .01$. Post hoc Bonferroni tests (.01) revealed that Descriptions aided identification significantly more than either Composites alone or Descriptions and Composites combined. The difference between Composites alone and Descriptions and Composites combined was not significant. No interaction was found to be significant.

9.2. Analysis of composites' ratings

9.2.1. Analysis based on Correct Identifications

There were significant main effects of Identification Information, $F(2,32)=12.35$, $p<.01$, and Target Race $F(1,32)=10.64$, $p<.01$, with higher Likelihood of identification based on the information provided by White ($M=2.93$, $SD=0.78$) rather than Black ($M=2.43$, $SD=0.63$) participants. Bonferroni pairwise comparisons indicated the highest Likelihood for Descriptions ($M=3.24$, $SD=.-69$). The difference in Likelihood for Composites ($M=2.29$, $SD=0.77$) and Descriptions & Composites ($M=2.39$, $SD=0.41$) was non-significant. None of the interactions were found to be significant.

Finally, an ANOVA was performed on Confidence ratings for trials where the target was correctly identified. However, due to the lower number of trials where faces were correctly detected in the array, participant scores were collapsed across Race of Compositor and Race of Target Face. There was a significant main effect of Identity information, $F(2,48) = 5.19$, $p < .01$. Post hoc Bonferroni tests (.05) showed similar results to previous findings; participants had a higher confidence in choices based on the information provided in Descriptions ($M=3.15$) rather than in Descriptions plus Composites ($M=2.52$) and solely in Composites ($M=2.49$). A difference in Confidence between Composites and Descriptions plus Composites was not significant.

9.2.2. Analysis based on False Identifications

The main effects of Identification Information, $F(2,58)=8.94$, $p<.01$, and Compositor Race $F(1,29)=10.91$, $p<.01$, were found to be significant for the subjective level of Likelihood as dependent variable in the analysis between participants. There was a significant superiority of Descriptions ($M=3.15$) over Composites ($M=2.61$) and marginally significant over Descriptions & Composites ($M=2.76$). Bonferroni pairwise comparison further indicated that the difference between Composites and Descriptions & Composites was not significant. Participants declared higher Likelihood of identification based on the information provided by White ($M=2.91$) rather than by Black ($M=2.77$) witnesses.

There was also a significant interaction Identification Information, X Compositor Race $F(2,58)=3.88$, $p<.05$ (See Figure 4). Identification Information provided in Descriptions by White ($M=3.30$) participants was rated higher on Likelihood scale than the information given by Black ($M=3.01$) witnesses. The differences among Information provided by participants from different ethnic backgrounds were not significant for other sources of Identification Information. The Descriptions were rated to be significantly more likely to lead to identification rather than Composites ($M=2.6$) and Descriptions & Composites ($M=2.82$)

provided by White witnesses. Similar findings were found for information provided by Black participants, with Descriptions being superior in Likelihood ratings (M=2.62 for Composites and M=2.69 for Descriptions & Composites). The differences between Composites and Descriptions & Composites were insignificant both for White and Black participants.

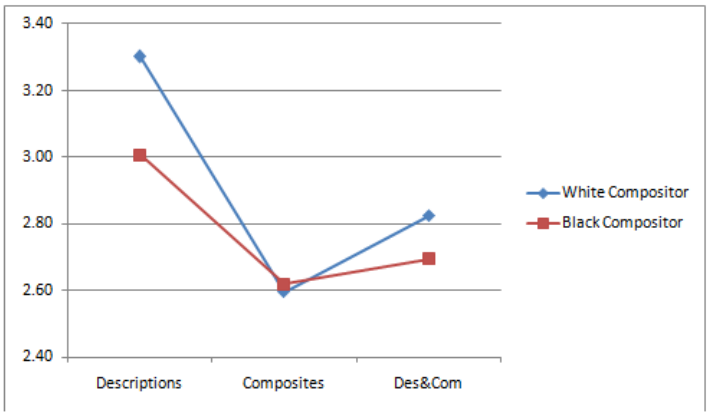

Figure 4 A significant interaction Identification Information X Compositor Race for Likelihood as dependent variable in Misidentification condition

Finally, an ANOVA was performed on Confidence in False Identifications. There was a significant main effect of Identification Information $F(2,58)= 7.28$, $p<.01$. Like in previous analysis (Bonferroni), there was a superiority of Descriptions (M=2.88, SD=0.27) over Descriptions & Composites (M=2.63, SD=0.22) and Composites (M=2.52, SD=0.25). A difference between Descriptions & Composites and Composites did not reach a significant level.

Analysis yielded also a significant main effect of Race of Compositor $F(1,29)= 13.07$, $p<.001$. Participants were more confident in identification about information provided by White compositors (M=2.72) rather than Black compositors (M=2.58). No other main effects or interactions reached significance.

10. Discussion

Analysis revealed that Descriptions aided identification significantly more than either Composites alone or Descriptions and Composites combined. The results reported here replicate previous studies demonstrating that the descriptions lead to more accurate identifications than composites [13]. Providing detailed descriptions of a person is a very difficult task which requires a person to put into words a perceptual experience [19]. The most typical technique of extracting the descriptions from witness is simply asking him/her to list features describing the offender's face (feature-based). However, a number of studies have reported a beneficial role of a more holistic approach to collecting descriptions, where a person is asked to describe a suspect by giving details about the kind of person she/he looked like i.e. good/bad [20]. Constructing a facial composite is preceded by giving a verbal description. Therefore, one would expect better results from a holistic system while using a holistic description. In the current study, a person provided a feature-base description and just after that she/he started working on the composites using holistic software. One might argue that in such situation witness experience "high cognitive load", firstly to change a visual experience into a verbal one and then verbal into visual again. Taking this under

consideration creating a composite can lead to memory demands which exceed recall abilities; this might explain the lower accuracy gained from Composites rather than from Descriptions.

The evaluation of confidence ratings across the different identity information conditions shows a pattern complementary to identification accuracy. However, even when participants make an incorrect decision they are still more confident if it had been based on a description rather than when it is based on a composite and confidence is lowered when the description is presented in combination with the composite. It is proposed that it is the weaker information available from the composite that produces this decrease in confidence. An examination of confidence level for trials where the target face was correctly identified confirm the same pattern, Participants rated their confidence as being higher when it was based on the description alone rather than on a composite or the description and composite combined. Pairing the description with a composite appears to lower the confidence that the description inspires on its own.

It seems surprising that the lower accuracy was found for identifications based on the Descriptions & Composites. One would possibly expect at least a similar level of accuracy between Descriptions and Descriptions & Composites since the second one is partially based on the same information provided in Descriptions, which are found to be the dominant source of information. However, the literature shows that faces are 'prone to grab one's attention' [21], which might explain why composites seem to play a dominant role in identification procedure.

Finally, as it was mentioned previously the process of matching unfamiliar faces relies on image matching rather than on a more sophisticated face matching strategy [14]. Therefore, one is comparing the individual features between the target photograph and the composite which can never be exactly the same, rather than overall construct of the composite and target [15]. One can expect that participants will never be fully satisfied with the composites, simply because the composites will never be identical with the target face [22].

11. General Discussion

Christie and Ellis [9] suggested that their finding that facial composites yielded poorer identification than a verbal description was due to the way that composites were created. Older systems like the one they employed, Photofit, involve construction on a feature-by-feature basis. This process may be incompatible with the form of representation humans typically use to store memories of faces. If this were the case then we may have expected that a facial composite derived from a system employing a holistic basis such as EFIT-V would not suffer in comparison to descriptions. However, the findings from the two experiments reported here show no evidence to support that. In Experiment 1 subjective evaluation of a piece of information in relation to its likely utility in leading to the identification of a target showed an advantage for descriptions over composites. Furthermore, this advantage for the description remained even when the composite was presented in conjunction with the description. In Experiment 2 the same outcome is demonstrated when the different types of information are given to participants to aid them in an identification task. Furnishing descriptions helped this process more than furnishing a composite. In addition, the description alone led to superior performance than presenting the description combined with the composite.

These findings provide support for the police practice of only circulating a small subset of what they presumably judge to be the best composites. However, it remains necessary to find an explanation for why composites lead to such poor identification performance.

Undoubtedly, providing detailed descriptions of a person is a very difficult task that requires a person to put into words a perceptual experience [19]. Although, both this study and that of Christie and Ellis have found that descriptions lead to better identification than composites the absolute levels of identification are low in both cases. Nevertheless, the real question of interest is why composites appear to be such a poor way of externalising the internal representation of a briefly seen face. Although it is worth noting that the internal representation may be itself very poor, the task in relation to forensic identification is still one of finding ways to improve performance with facial composites.

An answer to this question might be constructed by considering two curious findings. First, it is interesting to note the lower accuracy in identifications based on the Descriptions plus Composites. One would possibly expect at least a similar level of accuracy between Descriptions and Descriptions plus Composites since the second one contains all the same information provided in Descriptions, which are found to be the dominant source of information. However, this difference was also seen in the rated evaluations collected in Experiment 1 (and in a non-significant difference in the data from Christie and Ellis [9]. How might this arise? It is known that faces are 'prone to grab one's attention' [21]. A human face is a very significant stimulus for human beings because it carries many important social and biological cues. Therefore, a human face, and in the current research the high quality of the composites makes them equivalent to faces, is very hard to ignore. Consequently, a face always takes priority regardless of task relevance and attention load. Due to the importance of sociobiological cues carried in the faces; these cannot be ignored and are automatically and mandatorily processed. Therefore, the effect of lowered accuracy of identifications based on the Identification Information provided in the Descriptions plus Composites in comparison to information provided solely in Descriptions might be to an extent explained by Composites acting as distractors. Participants in the current study simply could not ignore the Composites, even if they were providing less valuable cues of correct identifications than accompanying Descriptions.

A second clue to understanding this effect comes from a consideration of the confidence ratings in Experiment 2 expressed on trials where the wrong target face was selected. Even though participants are making an error they are still more confident of their choice if it were based on a description. This suggests that in some fashion the information in the description is seen as more compelling than the information present in a composite.

Older composite systems like Photo-Fit or Identi-Kit have been criticised for providing poor quality of facial composites [22] which could have explained the better utility of verbal descriptions than the composites found by Christie and Ellis [9]. However, as can be seen from Figure 1b the composites produced using EFIT-V are extremely face like in appearance. We suggest that is unlikely that any further advances in technology are likely to significantly improve the quality of composites themselves. We suspect that part of the problem may now arise from the excellence of the representation constructed. The composites are so face like that participants in studies like this one expect it to be identical with the face of offender/target, yet the constructor can only reach as good a likeness as their internal memory representation allows.

In summary, we propose that further improvement in the identification performance achieved with facial composites must come from changes in the way that they are elicited from witnesses. Given that being a witness is likely to be at most a once in a lifetime experience, it is essential to invest in the way that composite system operators elicit and interpret the information provided by witnesses. In addition, in experimental settings both descriptions and composites are highly criticised due to the impairment of the following identification they can cause. However, the main tasks of the composites and descriptions was

too narrow down the pool of potential suspects and to allow identification of an offender by someone from the public. Therefore it is suggested to publicise both sources of the information together; as it seems that they make a significant contribution to the identification. It is likely that descriptions and composites complete each other and overcome the limitations each of them posses separately.

References

[1] Johnston, R.A., & Edmonds, A. "Recognising familiar and unfamiliar faces" *Memory, 17*, 2009, pp. 577-596
[2] Brace, N. A., Pike, G. E., Allen, P., & Kemp, R. I. "Identifying composites of famous faces: Investigating memory, language and system issues" *Psychology, Crime and Law12*(4), 2006, pp. 351-366.
[3] Davies, G. M., & Christie, D. Face Recall: An Examination of Some Factors Limiting Composite Production Accuracy. *Journal of Applied Psychology, 67*(1), 1982, 103-109.
[4] Frowd, C., Carson, D., Ness, H., McQuiston-Surrett, D., Richardson, J., Baldwin, H., et al. Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology,10*,2005, 63-81.
[5] Frowd, C., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S., et al. (). A forensically valid comparison of facial composite systems. *Psychology, Crime and Law, 11*(1), 2005, 33-52.
[6] Frowd, C., McQuiston-Surrett, D., Kirkland, I., & Hancock, P. The process of facial composite production. In A. Czerederecka, T. Jaskiewicz-Obydzinska, R. Roesch & J. Wojcikiwiecz (Eds.), *Forensic Psychology and Law*. 2004, Krakow: Institute of Forensic Research Publishers.
[7] Tredoux, C. G., Meissner, C. A., Malpass, R. S., & Zimmerman, L. A. "Eyewitness Identification". *Encyclopedia of Applied Psychology, 4,* 2004
[8] Brown, C., Lloyd-Jones, T. J., & Robinson, M. "Eliciting person descriptions from eyewitnesses: A survey of police perceptions of eyewitness performance and reported use of interview techniques" *European Journal of Cognitive Psychology, 20*(3), 2008, pp. 529-560.
[9] Christie, D. F. M., & Ellis, H. D. "Photofit Constructions Versus Verbal Descriptions of Faces" *Journal of Applied Psychology, 66*(3), 1981, pp. 358 - 363.
[10] Davies, G. M. & Valentine, T. Facial composites: forensic utility and psychological research. In: R.C.L. Linsay, D.F. Ross, J.D. Read & M. P. Toglia *Handbook of eyewitness psychology. Volume 2: Memory for people.* Mahwah: LEA., 2007, pp. 59-83
[11] Gibson, S.J., Solomon, C.J., Pallares-Bejarano, A. Synthesis of photographic quality facial composites using evolutionary algorithms. *Proceedings of the British Machine Vision Conference*, R. Harvey and J.A. Bangham, (Eds), 2004, pp. 221-230
[12] Frowd, C. D., Hancock, P.J.B. & Carson, D. "EvoFIT: A holistic, evolutionary facial-imaging technique for creating composites. *ACM Transactions on Applied Psychology (TAP), 1*, 2004, pp. 1 – 27
[13] Frowd, C., Pitchford , M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., et al. "The psychology of face construction: giving evolution a helping hand" *Applied Cognitive Psychology, 25,* 2010
[14] Megreya, A. M., & Burton, A. M. "Hits and false positives in face matching: A familiarity-based dissociation" *Perception and Psychophysics, 69*(7), 2007, pp. 1175-1184.
[15] Schmidt, H. C., & Frowd, C.. *An Investigation into the Cross-Race Effect in Face Composite Construction.* Paper presented at the Proceedings of the 16th Conference of the European Association of Psychology and Law London, 2006
[16] McQuiston-Surrett, D., Topp, L. D., & Malpass, R. S. "Use of Facial Composite Systems in US Law Enforcement Agencies" *Psychology, Crime and Law, 12*(5), 2006, pp. 505-517.
[17] Tanaka, J. W., & Sengco, J. A. "Features and their configuration in face recognition". *Memory and Cognition, 25*(5), 1997, pp. 583-592.
[18] Pezdek, K., Blandon-Gitlin, I., & Moore, C. "Children's Face Recognition Memory: More Evidence for the Cross-Race Effect" *Journal of Applied Psychology, 88*(4), 2003, pp. 760-763.
[19] MacLin, M. K. "The effects of exemplar and prototype descriptors on verbal overshadowing" *Applied Cognitive Psychology, 16*, 2002, pp. 929-936.
[20] Wickham, L. H. V., & Lander, K. "The effect of verbal description and processing type on face identification" *European Journal of Cognitive Psychology, 20*(3), 2008, pp. 577-586.
[21] Brédart, S., Delchambre, M., & Laureys, S."One's Own Face Is Hard to Ignore." *Quarterly Journal of Experimental Psychology, 59*(1), 2006, pp. 46-52.
[22] Koehn, C. E., & Fisher, R. P. "Constructing Facial Composites with the Mac-A-Mug Pro System" *Psychology, Crime and Law, 3*, 1997, pp. 209-218.